Titre: Title:	Automated Segmentation of the Spinal Cord on Magnetic Resonance Echo Planar Images
Auteur: Author:	Rohan Banerjee
Date:	2025
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Banerjee, R. (2025). Automated Segmentation of the Spinal Cord on Magnetic Resonance Echo Planar Images [Mémoire de maîtrise, Polytechnique Montréal]. PolyPublie. https://publications.polymtl.ca/65808/

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/65808/
Directeurs de recherche: Advisors:	Benjamin De Leener, & Julien Cohen-Adad
Programme: Program:	Génie informatique

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Automated Segmentation of the Spinal Cord on Magnetic Resonance Echo Planar Images

ROHAN BANERJEE
Département de génie informatique et génie logiciel

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
Génie informatique

Mai 2025

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Ce mémoire intitulé :

Automated Segmentation of the Spinal Cord on Magnetic Resonance Echo Planar Images

présenté par **Rohan BANERJEE**
en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
a été dûment accepté par le jury d'examen constitué de :

Nikola STIKOV, président
Benjamin DE LEENER, membre et directeur de recherche
Julien COHEN-ADAD, membre et codirecteur de recherche
Hérve LOMBAERT, membre

DEDICATION

I dedicate this thesis to my mother Balaka Banerjee
and my brother Arjun Banerjee.

ACKNOWLEDGEMENTS

This master's journey has been a profound learning experience, shaping not only my understanding of research but also enriching multiple facets of my life immeasurably. It's a path I did not walk alone, and I am immensely grateful for the incredible community of mentors, colleagues, and friends whose support, wisdom, and encouragement have been the bedrock upon which this work stands.

First and foremost, my deepest gratitude goes to my advisors, Prof. Benjamin De Leener and Prof. Julien Cohen-Adad. Thank you for your unwavering support and expert guidance as I navigated the complex and exciting terrain of my master's research endeavours. You opened the door to this whole new area of research for me, and I am profoundly thankful for the opportunity. I would also like to extend my sincere thanks to my close collaborators, Dr. Merve Kaptan and Prof. Kenneth Weber from Stanford University; your continuous support and insightful feedback were crucial throughout this project.

My appreciation extends to the amazing folks at NeuroPoly. Jan Valosek, Naga Karthik, Sandrine Bédard, Andjela Dimitrijevic, Alexandre D'Astous, Pierre-Louis Benveniste, Nathan Molinier, Joshua Newton, Mathieu Guay-Paquet, Nick Guenther, and the entire lab community at large – thank you for creating such a welcoming and stimulating environment. Your willingness to patiently listen to my doubts and collaboratively help me solve them was invaluable. The lively discussions in the lab, the critical thinking fostered during reading group meetings, and the shared spirit of inquiry were not just beneficial but essential for carrying out this research.

Much of this journey also unfolded within the vibrant ecosystem of Mila – Quebec AI Institute, a place where I spent countless days and nights immersed in learning. Mila was far more than just a research workspace; it was a dynamic confluence of strong and brilliant minds from vastly different backgrounds and cultures, individuals exceptionally talented not only in AI but possessing a fascinating array of other skills alongside profound research insights. I feel incredibly fortunate to have landed in such a uniquely supportive, collaborative, and genuinely fun community. That this remarkable institution resides in the heart of Montreal was truly the cherry on the cake, making this experience all the richer.

Along this demanding yet rewarding path, I forged friendships that have become pillars of support, friendships with people I now cherish as family. Nikita Saxena, Arkil Patel, Nithya Shikarpur, Vedant Shah, Moksh Jain, Aniket Didolkar, Vineet Jain, Nanda Krishna, Jith Subramanian, Nishka Katoch, Shruti Joshi, Shivakanth Sujit, Prishruit Punia, Divyat

Mahajan, Rohan Sukumaran, Megh Thakkar, Aayush Bajaj, Shravan Nayak, Prashanth Govindrajan, Tejas Kasetty, Vishal Gupta and Tejas Vaidhya, thank you for the unwavering support, laughter, shared meals, late-night study sessions and the love.

My growth during this period was also significantly shaped by endeavours beyond the direct scope of this thesis, which were seminal in building my confidence and broadening my research perspective. I would particularly like to acknowledge Prof. Gauthier Gidel and Prof. Eilif Muller for the extraordinary opportunity to participate as an AI researcher at the Co-Creating AI and Art Residency. It provided a unique and powerful medium to express myself creatively through art, fostering collaborations with a diverse group of established artists and AI researchers – fulfilling a long-held aspiration. This experience profoundly underscored the importance and beauty of co-creation across disciplines.

Last, and most importantly, I would like to thank Maa and Dada – my mother and my elder brother – for making me who I am today. You both have defined patience and courage for me. Maa, you were my first teacher, imparting not only the logic of maths but, more importantly, the enduring value of kindness and patience, especially during the most difficult times. Dada, thank you for being my hero. Your lessons of perseverance and strength guide me literally every single day. Thank you both for your immense support in all my decisions and for always believing in me. You keep me going, and I am forever grateful to you.

To everyone who contributed to this journey, in ways seen and unseen, thank you from the bottom of my heart. This experience reaffirms that meaningful progress is rarely a solitary pursuit, but a testament to the power of community, collaboration, and the shared passion for discovery.

RÉSUMÉ

L'imagerie par résonance magnétique fonctionnelle (IRMf) de la moelle épinière (ME) offre des perspectives uniques sur le traitement sensoriel, moteur et autonome, présentant un potentiel significatif pour le suivi de la progression des maladies et de l'efficacité des traitements dans les affections neurologiques. Cependant, les progrès de l'IRMf de la ME ont pris du retard par rapport à l'imagerie cérébrale, entravés par des défis techniques tels que les artéfacts de mouvement, le bruit physiologique, les distorsions induites par la susceptibilité inhérentes à l'imagerie écho-planaire (EPI), et le manque d'outils de traitement standardisés. Une étape critique pour permettre une analyse de groupe reproductible est la normalisation spatiale précise des données d'IRMf de la ME individuelles à un espace de gabarit commun, qui repose fondamentalement sur une segmentation précise de la ME. Les méthodes de segmentation automatisée actuelles échouent souvent sur les données EPI de faible résolution et sujettes aux artéfacts typiquement utilisées pour l'IRMf, obligeant à recourir à la segmentation manuelle, qui est longue, laborieuse et subjective, ce qui introduit des biais et limite le passage à l'échelle.

Cette thèse présente EPISeg, une nouvelle méthode basée sur l'apprentissage profond spécifiquement développée pour la segmentation automatisée de la ME directement sur les images EPI à écho de gradient. Pour entraîner et valider rigoureusement EPISeg, nous avons constitué et rendu public un jeu de données EPI de la ME multi-sites à grande échelle (N=406 participants, 15 sites) avec des segmentations de référence vérifiées par des experts, organisé selon le standard BIDS (Brain Imaging Data Structure) et partagé via OpenNeuro. EPISeg, basé sur une architecture U-Net 3D, a été entraîné sur ce jeu de données diversifié en utilisant une stratégie itérative d'apprentissage actif avec humain dans la boucle (HIL-AL). Cette approche a optimisé l'utilisation des ressources limitées d'annotation par des experts tout en améliorant la robustesse du modèle face aux artéfacts d'imagerie courants et aux variations entre les différents protocoles d'acquisition.

Les évaluations démontrent qu'EPISeg surpasse les méthodes de segmentation de la ME de pointe existantes telles que PropSeg, DeepSeg et les méthodes de segmentation agnostique au contraste sur des données de test non utilisées pour l'entraînement, démontrant une grande précision (score de Dice = 0.88 ± 0.05) et une fidélité géométrique (distance de Hausdorff = 1.41 ± 0.86) sur des jeux de données diversifiés, incluant des volontaires sains et des populations de patients. En fournissant une solution fiable, automatisée et accessible au public, intégrée dans la Spinal Cord Toolbox (SCT), EPISeg lève un verrou majeur dans l'analyse de l'IRMf de la ME. Ce travail facilite une recherche plus reproductible, efficace et réalisable à plus grande échelle, faisant ainsi progresser notre capacité à étudier la fonction de la ME humaine chez le sujet sain et en pathologie.

ABSTRACT

Functional magnetic resonance imaging (fMRI) of the spinal cord (SC) offers unique insights into sensory, motor, and autonomic processing, holding significant potential for monitoring disease progression and treatment efficacy in neurological conditions. However, the advancement of SC fMRI has lagged behind brain imaging, hampered by technical challenges including motion artifacts, physiological noise, susceptibility-induced distortions inherent in echo-planar imaging (EPI), and a lack of standardized processing tools. A critical step for enabling reproducible group-level analysis is the accurate spatial normalization of individual SC fMRI data to a common template space, which fundamentally relies on precise SC segmentation. Current automated segmentation methods often fail on the low-resolution, artifact-prone EPI data typically used for fMRI, forcing reliance on time-consuming, labor-intensive, and subjective manual segmentation, which introduces bias and limits scalability.

This thesis introduces EPISeg, a novel deep learning-based method specifically developed for the automated segmentation of the SC directly on gradient-echo EPI images. To train and rigorously validate EPISeg, we curated and publicly released a large-scale, multi-center SC EPI dataset (N=406 participants, 15 sites) with expert-verified ground-truth segmentations, organized according to the Brain Imaging Data Structure (BIDS) standard and shared via OpenNeuro. EPISeg, leveraging a 3D U-Net architecture, was trained on this diverse dataset using an iterative human-in-the-loop active learning (HIL-AL) strategy. This approach optimized the use of limited expert annotation resources while enhancing model robustness against common imaging artifacts and variations across different acquisition protocols.

Evaluations demonstrate that EPISeg outperforms existing state-of-the-art SC segmentation tools like PropSeg, DeepSeg and Contrast-agnostic segmentation methods on held-out test data, exhibiting high accuracy (Dice score=0.88 ± 0.05) and geometric fidelity (Hausdorff distance=1.41 ± 0.86) across diverse datasets, including healthy volunteers and patient populations. By providing a reliable, automated, and publicly accessible solution which is integrated into the Spinal Cord Toolbox (SCT), EPISeg addresses a major bottleneck in SC fMRI analysis. This work facilitates more reproducible, efficient, and scalable research, ultimately advancing our ability to investigate human spinal cord function in health and disease.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ACRONYMS

AL	Active Learning
BIDS	Brain Imaging Data Structure
BOLD	Blood Oxygentation Level Dependent
CNN	Convolutional Neural Network
CSF	Cerebrospinal fluid
DL	Deep Learning
DS	Dice Score
EPI	Echo Planar Imaging
FA	Flip Angle
FCN	Fully Convolutional Networks
fMRI	Functional Magnetic Resonance Imaging
FoV	Field of View
FSE	Fast Spin Echo
GE	Gradient-Echo
HD	Hausdorff Distance
HD95	95th percentile Hausdorff Distance
HIL-AL	Human-in-the-loop Active Learning
HRF	Haemodynamic Response Function
HV	Healthy Volunteers
JSON	JavaScript Object Notation
MIS	Medical image segmentation
MRI	Magnetic Resonance Imaging
PD	Proton Density
PVE	Partial volume effect
QC	Quality Control
ReLU	Rectified Linear Unit
SC	Spinal Cord
SCT	Spinal Cord Toolbox
SE	Spin Echo
TE	Echo Time
TR	Repetition Time
TSV	Tab-Separated Values

CHAPTER 1 INTRODUCTION

Magnetic resonance imaging (MRI) has revolutionized our ability to non-invasively explore the structure and function of the central nervous system (CNS), including both the spinal cord (SC) and brain. SC MRI, in particular, has provided critical insights into the neuroanatomy of the cord and has become essential for diagnosing and monitoring neurological disorders such as multiple sclerosis, SC injury, and neurodegenerative diseases [9]. Over the past decades, advances in hardware and image processing techniques have enabled increasingly detailed investigations of SC morphology.

Building on structural insights, functional magnetic resonance imaging (fMRI) has emerged as a valuable technique for non-invasively investigating SC function in both human and animal models [8, 10]. This modality enables researchers to map neuronal activation associated with sensory, motor, and autonomic processes. For instance, studies have utilized spinal fMRI to observe activity in response to various intensities of touch and pressure stimuli, demonstrating its capability to detect nuanced SC activity [11]. Additionally, spinal fMRI has been employed to study pain processing and its descending modulation, providing insights into spinal nociceptive mechanisms and individual differences [12]. Moreover, spinal fMRI shows promise in the context of SC injury for assessing residual neural activity and functional changes post-injury [13].

Despite this potential, the application and advancement of SC fMRI have significantly lagged behind brain fMRI research, reflected in the comparatively smaller body of published literature [8]. This disparity stems largely from substantial technical challenges inherent to imaging the SC. Anatomical constraints, including the SC's small diameter and elongated structure, necessitate a difficult trade-off between spatial resolution (to minimize partial volume effects (PVE)) and signal-to-noise ratio (SNR), often requiring extended fields of view. Furthermore, the SC's location within the vertebral column, surrounded by tissues with differing magnetic properties (bone, CSF, discs, air), leads to significant magnetic field inhomogeneities. These result in susceptibility artifacts, manifesting as geometric distortions and signal dropouts, particularly problematic for common fMRI sequences [14]. Compounding these issues is physiological noise; the proximity to the heart and lungs induces motion artifacts and pulsatile CSF flow linked to cardiac and respiratory cycles, further degrading image quality and complicating signal interpretation [15].

These challenges are particularly acute for the gradient-echo (GE) echo-planar imaging (EPI) sequences most commonly used in fMRI due to their speed and sensitivity to the BOLD effect. EPI's inherent sensitivity to magnetic field inhomogeneities drastically exacerbates susceptibility artifacts in the SC region, leading to pronounced image distortion and signal loss [14]. The combination of these artifacts, physiological noise, and the typically low spatial resolution of fMRI EPI scans severely compromises data quality and hinders the application of standard fMRI analysis pipelines.

A critical step impeded by poor EPI data quality is spatial normalization which is the process of aligning individual subject data into a common anatomical space to enable group-level statistical analysis and comparison across studies. Robust normalization relies heavily on accurate segmentation of the anatomical structure of interest - SC in our case. However, segmenting the SC directly on the distorted, low-contrast EPI images is difficult. Existing automated segmentation tools, often designed for higher-quality structural MRI, typically fail or perform poorly on SC EPI data [16]. Consequently, researchers often resort to laborious manual segmentation or correction, a process that is not only time-consuming but also introduces significant subjectivity and inter-rater variability, undermining the goal of reproducible research [17]. This lack of a reliable automated segmentation method for SC EPI data represents a major bottleneck limiting the scalability and reproducibility of SC fMRI studies.

The primary objective of this thesis is to address this critical gap by developing and validating a robust, automated method for segmenting the SC directly on GE EPI data. We introduce EPISeg, a deep learning (DL) model based on the 3D U-Net architecture, specifically designed for this task. Recognizing the challenge of obtaining large annotated datasets, EPISeg was trained using an iterative human-in-the-loop active learning (HIL-AL) strategy. The contributions of this works are: 1) The creation and public release of a large, multi-center, open-access SC EPI dataset with expert-verified segmentations to facilitate model development and validation. 2) The development, rigorous evaluation, and public release via integration into the Spinal Cord Toolbox (SCT) of the EPISeg segmentation tool.

The remainder of this thesis is organized as follows: Chapter 2 provides a comprehensive literature review covering the foundational medical concepts related to the SC and MRI/fMRI, as well as the relevant DL principles, equipping the reader with the necessary background for the subsequent chapters. Chapter 3 details the overall methodology, breaking down the specific contributions of this thesis. Chapter 4 presents the core technical work, outlining the motivation, experimental design, implementation, and results for the proposed EPISeg method. In Chapter 5 we present a discussion. Finally, Chapter 6 concludes the thesis, summarizing the key findings, and offers perspectives on future work.

CHAPTER 2 LITERATURE REVIEW

2.1 Medical Review

This section focuses on SC imaging, specifically exploring the core concepts, applications, and challenges associated with MRI and fMRI. While not exhaustive, the goal is to establish the necessary context and motivation for the subsequent chapters by familiarizing the reader with the principles and limitations of current SC imaging techniques.

2.1.1 Spinal Cord

The SC, a vital component of the CNS, serves as the primary conduit for neural signals between the brain and the peripheral nervous system. This elongated, cylindrical structure which is encased within a protective vertebral column, extends from the medulla oblongata in the brainstem to approximately the L1-L2 vertebral level in adults, where it tapers into the conus medullaris and subsequently the cauda equina, a collection of nerve roots that continue caudally [18]. Understanding the intricate anatomy and physiology of the SC is paramount, as it plays a critical role in mediating sensory perception, motor control, and autonomic functions essential for human life [19].

The SC is organized into distinct segments – cervical, thoracic, lumbar, sacral, and coccygeal – each giving rise to pairs of spinal nerves that innervate specific regions of the body. Internally, as seen in Fig. 2.1, it exhibits a characteristic butterfly-shaped grey matter core which is surrounded by white matter. The grey matter primarily consists of neuronal cell bodies, dendrites, and synapses, and is further divided into dorsal horns (receiving sensory information), ventral horns (containing motor neurons), and intermediate horns (involved in autonomic functions). In contrast, the white matter is composed of myelinated axons organized into ascending (sensory) and descending (motor) tracts that facilitate rapid communication over long distances within the CNS. Key tracts such as the corticospinal tract (motor control), spinothalamic tract (pain and temperature), and dorsal columns (touch and proprioception) are crucial for integrated bodily function. Furthermore, the entire SC is enveloped by three protective layers of meninges – the dura mater, arachnoid mater, and pia mater – which provide physical cushioning and contribute to the blood-brain barrier, safeguarding the delicate neural tissue [20].

A wide range of debilitating impairments, including paralysis, sensory loss and autonomic dysregulation can be a result of dysfunction of the SC, whether due to traumatic injury, or

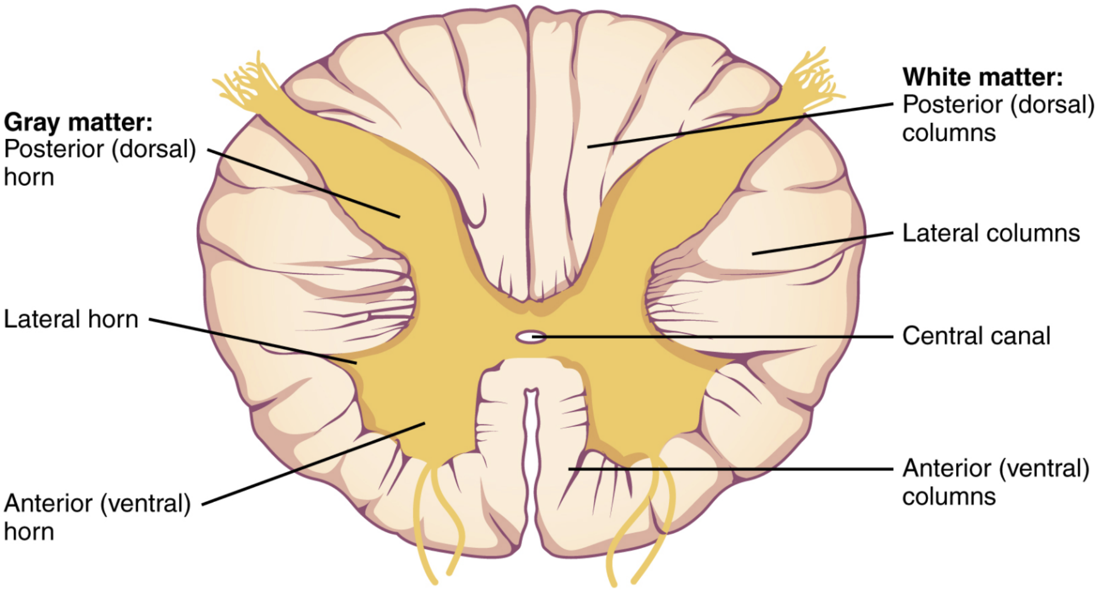

Figure 2.1 Anatomical cross-section of the spinal cord. Source: OpenStax Anatomy and Physiology, CC-BY license.

degenerative conditions [21]. Therefore, non-invasive ways for studying SC structure and function such as MRI are of significant clinical and research interest. MRI allows for detailed visualization of the SC's anatomy which enables the identification of pathological changes and provides insights into the neural mechanisms. Furthermore, advancements in fMRI techniques help for investigating the dynamic neural activity within the SC in vivo and offers opportunities to understand complex human behaviors and neurological disorders [19].

2.1.2 Magnetic Resonance Imaging

Acquisition

MRI leverages the magnetic properties of atomic nuclei, primarily hydrogen protons abundant in water and fat, to generate signals. The basic principle involves placing the subject within a strong static magnetic field (B_0), which aligns the protons' magnetic moments. Radiofrequency (RF) pulses are then applied to perturb this alignment, and as the protons return to their equilibrium state, they emit detectable RF signals [12]. By manipulating the timing and characteristics of these RF pulses and applying magnetic field gradients, different contrasts can be achieved, enabling the differentiation of various tissue types based on their relaxation properties [22].

The commonly utilized MRI contrasts for anatomical imaging are T1-weighted, T2-weighted, and Proton Density (PD)-weighted images. These contrasts are primarily achieved using pulse sequences like Spin Echo (SE) and Fast Spin Echo (FSE). The T1-weighted imaging is sensitive to T1 relaxation time (TR) which is particularly effective in differentiating tissues based on their fat and water content. T2-weighted imaging on the other hand is sensitive to the T2 relaxation time which is highly sensitive to water content and edema, making it crucial for visualizing fluid-filled spaces like the cerebrospinal fluid (CSF) and detecting pathological fluid accumulation [22]. Another important contrast is generated using EPI, a rapid acquisition technique particularly crucial for fMRI due to its speed in capturing dynamic signal changes. EPI sequences are typically T2*-weighted and sensitive to BOLD contrast, which forms the basis of fMRI [23, 24].

Spatial encoding in MRI is achieved through the application of linear magnetic field gradients in three orthogonal directions (x, y, and z) during signal acquisition. These gradients induce spatially varying magnetic fields, causing protons at different locations to precess at slightly different frequencies. The spatial origin of the emitted RF signals can be decoded by systematically varying these gradients. This captured raw data, which is acquired in k-space can be reconstructed into a spatially resolved using Fourier transform techniques [25]. The quality and characteristics of the acquired images are significantly influenced by the choice of acquisition parameters, including TR, TE, FA, and voxel size [22].

Challenges

SC imaging presents several inherent challenges that require careful consideration for reliable data acquisition and interpretation. A primary obstacle is the intrinsically low SNR, especially when high spatial resolution is desired to resolve the fine structures within the small SC. This limited SNR is exacerbated by the surrounding CSF and bony structures.

Physiological motion from respiration and cardiac pulsation remains a significant source of artifacts. These involuntary movements can lead to blurring and distortions which make it particularly detrimental for long acquisitions needed for functional imaging [26, 27]. Along with this, magnetic susceptibility artifacts also pose a considerable challenge. These artifacts arise induce field inhomogeneities causing geometric distortions and signal loss, arising from the tissue interfaces which have differing magnetic susceptibilities around the spine including bone, spine and CSF.

The PVE is another critical challenge in SC MRI. This effect arises because each image voxel represents a finite volume of tissue and a single voxel can contain multiple tissue types like grey matter, white matter, CSF etc, in regions with complex anatomy like the SC [28]. This averaging of signals from different tissues within a voxel can reduce the accuracy of quantitative measurements, blur boundaries between structures, and confound the interpretation of functional signals. Finally, practical limitations arise from the often lengthy acquisition times required for high-quality SC MRI. Longer scans increase the risk of subject motion and discomfort and potentially degrade the data quality.

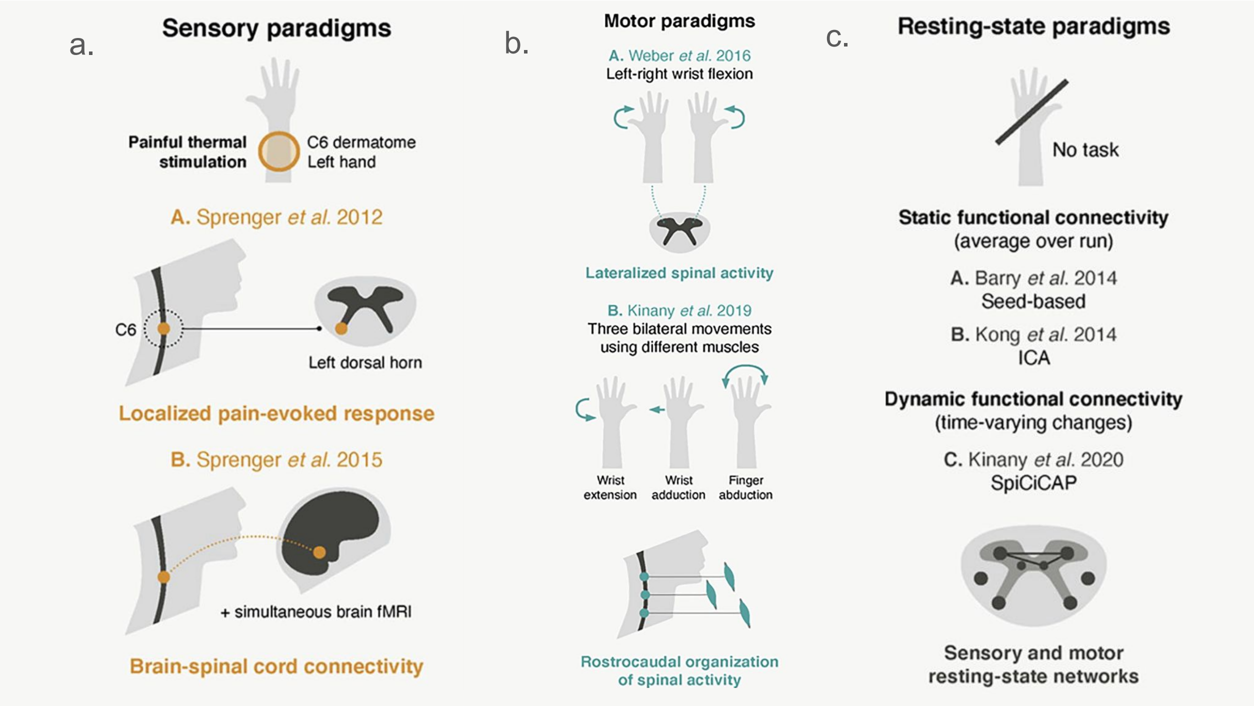

Figure 2.2 Spinal cord fMRI study paradigms. (a) Shows studies of spinal cord fMRI using sensory paradigm, (A) Sprenger C., et al [1] and (B) Sprenger C., et al [2]; (b) Shows studies of spinal cord fMRI using motor paradigm, (A) Weber K.A., et al [3] and (B) Kinany N., et al [4]; (c) Shows studies of spinal cord fMRI using resting-state paradigm (A) Barry R., et al [5], (B) Kong Y., et al [6], (C) Kinany N., et al [7]; Source: Figure adapted from Kinany N., et al, [8], CC-BY license.

2.1.3 Functional Magnetic Resonance Imaging

FMRI extends the capabilities of structural MRI to investigate brain and SC function by indirectly measuring neural activity through blood flow changes. Unlike anatomical MRI, which provides a static snapshot of tissue structure, fMRI captures dynamic changes in blood oxygenation levels that are coupled to neuronal activity, allowing for the study of brain and SC responses to various stimuli or tasks. The most widely used fMRI technique relies on the Blood-Oxygen-Level-Dependent (BOLD) contrast mechanism. Neural activity triggers a cascade of physiological events, including increased local cerebral blood flow (CBF) and oxygen consumption. Although oxygen consumption increases, the increase in CBF is typically proportionally larger, leading to a net decrease in the concentration of deoxyhemoglobin in the venous capillaries and venules in the active region. Deoxyhemoglobin is paramagnetic, meaning it distorts the local magnetic field, whereas oxyhemoglobin is diamagnetic and has minimal effect. This change in the ratio of deoxyhemoglobin to oxyhemoglobin alters the local magnetic susceptibility, which can be detected as a change in the T2* signal, forming the basis of the BOLD contrast. Therefore, increased neural activity indirectly leads to an increase in the fMRI signal [14].

The temporal dynamics of the BOLD response are characterized by the Haemodynamic Response Function (HRF). During the period of neural activity the BOLD signal exhibits an initial dip which is followed by a delayed and smoothed positive peak for around 4-6 seconds after stimulus onset and then returns to baseline. This HRF reflects the complex interplay of neurovascular coupling and is considered when designing fMRI experiments and interpreting results. fMRI experiments are typically designed using paradigms that manipulate cognitive, sensory, or motor processes to evoke changes in neural activity. Fig. 2.2 depicts a few examples of SC fMRI studies. Common experimental designs include block designs where periods of task and rest are alternated and event-related designs, where individual stimuli or events are presented in a randomized fashion. The acquired fMRI data then undergoes several preprocessing steps, including motion correction, slice-timing correction, spatial smoothing, and statistical analysis to identify brain or SC regions showing significant BOLD signal changes related to the experimental paradigm. fMRI has become an invaluable tool for mapping brain and, increasingly, SC function in vivo, providing insights into the neural basis of behavior, cognition, and neurological disorders [14].

2.1.4 Functional Magnetic Resonance Imaging in Spinal Cord

Applications of fMRI in SC is a more recent and technically demanding compared to the well-established techniques for studying brain function. The fundamental principles of BOLD fMRI, relying on neurovascular coupling to infer neuronal activity from hemodynamic changes are applicable to the SC. However, the SC presents unique anatomical and physiological challenges that necessitate significant adaptations of standard fMRI approaches used in the brain [8, 14].

As discusssed in the previous section, firstly, the cross-sectional area of the SC is significantly smaller than that of the brain which limits the number of signal-generating protons in each voxel. Therefore, the inherent SNR of SC is lower than the brain. As highlighted by [14], this

SNR limitation is a major hurdle, requiring specialized hardware and acquition strategies.

Secondly, the susceptibility artifacts are more pronounced in SC imaging. The bony structure and the air-filled lung near around the SC creates inhomogeneities in the main magnetic field (B_0). These inhomogeneities lead to image distortions and signal dropout, especially with the commonly used GE EPI sequences for fMRI.

Thirdly, physiological noise from cardiac and respiratory pulsations make SC fMRI highly susceptible to motion artifacts. These artifacts often exceed the magnitude of the BOLD signal itself which could obscure neural activity and require sophisticated mitigation strategies.

Beyond acquisition challenges, preprocessing and analyzing SC fMRI data also presents unique hurdles, particularly concerning spatial normalization and group-level analysis. As Hoggarth et al. [17] emphasize, while SC fMRI offers substantial potential for elucidating sensory and motor functions, as well as the intricate interactions between the brain and SC [19,29–31], robust group-level inferences necessitate the precise co-registration of data into a standardized common space. Unlike brain fMRI where automated registration pipelines are more readily available, pipelines for SC fMRI are less readily available. Our research group is actively developing the SCT [32], a software package designed to process multiparametric SC data. Even though it provides readily available tools for preprocessing and analysis of SC, SC fMRI often necessitates manual steps, such as manual masking of the SC for registration. This manual intervention introduces a potential source of variability and bias, as the precise delineation of the SC boundary can be subjective and influenced by image quality and rater experience. The work [17], specifically investigates the impact of variability in manually drawn SC masks on subsequent registration and interpretation of fMRI results, highlighting the need for standardized and potentially automated analysis workflows to improve the reliability and reproducibility of SC fMRI research.

2.2 Brain Imaging Data Structure

The Brain Imaging Data Structure (BIDS) is a widely adopted community standard for organizing and describing neuroimaging data, aiming to enhance data sharing, reproducibility, and the development of automated analysis pipelines [33]. In the context of increasingly complex neuroimaging studies and the growing emphasis on open science, BIDS provides a structured and standardized framework for data organization, metadata annotation, and provenance tracking [34]. BIDS principles ensure that the datasets are easily understandable, accessible, and resusable by the broader scientific community and hence facilitates collaboration and accelerates scientific discovery.

BIDS defines a hierarchical folder structure for organizing neuroimaging data by using consistent naming conventions for the files and directories. The naming conventions typically consist of subject names or ids, sessions and modalities, for instance, anat for anatomical scans, dwi for diffusion-weighted image scans and func for functional images scans. The formatting structure expects the users to provide the metadata files of the data in JSON (JavaScript Object Notation) and TSV (Tab-Separated Values). The metadata may include values like experimental contexts, acquisition parameters and subject demographics. It supports a wide range of modalities which include structural MRI, fMRI, diffusion MRI, electroencephalography - EEG, magnetoencephalography - MEG with ongoing efforts to accommodate new and emerging imaging techniques.

We found the benefits of adapting BIDS for our work in three folds. Firstly, BIDS promotes simple data organization and management which makes it easy to navigate and understand complex datasets, especially in longitudinal and/or multi-site studies. Secondly, it significantly enhances reproducibility and data sharing. It also facilitates replication of research findings. Thirdly, it helps in the development of automated analysis pipelines, as the standardized data structure allow for heuristic-based access and interpretation of the data, reducing manual data handling.

While BIDS was mainly developed focusing on brain imaging, its principles are highly adaptable for any anatomical structure and highly relevant for SC imaging. We used the BIDS standards to organize our SC fMRI data where we provide a multi-site SC fMRI (EPI sequence) data along with reliable SC segmentations. Platforms like OpenNeuro (`https://openneuro.org/`) have internal BIDS format checkers which makes sure that the uploaded datasets adhere to the the rules. We have made our BIDS data publicly available on OpenNeuro.

2.3 Deep Learning Review

The following sections focus on establishing context especially for DL related concepts. We first start with general overviews and then move on to specific concepts that would come handy to the readers in order to understand the concepts mentioned in 4.

2.3.1 Medical Image Segmentation

Medical image segmentation (MIS) is a fundamental task in medical image analysis that involves the partitioning of medical images into meaningful anatomical regions or structures of interest. Segmentation aims to delineate boundaries different tissue types, organs, and lesions or other clinically relevant entities and assign them labels. Accurate and reliable segmentation is a critical prerequisite for a wide range of applications, like quantitative analyses, computer-aided prognosis and diagnosis, surgical planning and much more.

The goal of MIS is to automate or semi-automate the process of segmentation and reduce subjectivity, time consumption associated with manual segmentation by clinicians. Manual segmentations, though considered gold standard in some contexts, is often labor-intensive, time-consuming, prone to inter-/intra-rater variability and not scalable to large datasets. Semi-automatic segmentation goes a step forward and requires some user inputs to initialize the segmentation process. It involves technqiues like region growing, level sets, or graph-based methods, aiming to reduce user effort while maintaining accuracy. The most autonomous method of segmentation is the automatic segmentation method. This segmentation method does not need user input, rather it uses properties from the data to carry out the segmentation. These methods often employ techniques based on thresholding, edge detection, clustering, deformable models, atlas-based segmentation, and, increasingly, DL-based approaches.

2.3.2 Medical Image Segmentation in Spinal Cord

MIS of the SC is of great importance for automated interpretation and quantitative morphometric analysis of SC MRI data. Accurate and reliable SC segmentation is not merely a preliminary step, but a crucial enabler for a multitude of downstream analyses and clinical applications. Firstly, it provides the foundation for automated and objective interpretation of SC MRI, reducing the reliance on subjective visual assessment and enabling faster, more consistent analysis, particularly vital in high-throughput research and clinical settings [35–45]. Secondly, SC segmentation is essential for morphometric analysis, allowing for the precise quantification of SC volume, cross-sectional area, and shape, which serve as critical biomarkers in the diagnosis, prognosis, and monitoring of various SC pathologies, including SC injury and neurodegenerative diseases [35,46]. Furthermore, accurate segmentation of the SC itself forms the basis for the subsequent segmentation of other crucial SC anatomies, such as the grey matter and white matter, detection of the SC centerline for accurate spatial referencing, identification and delineation of SC injuries, and the precise localization and volumetric assessment of SC lesions. Beyond single time-point analysis, robust SC segmentation has great potential for automating and facilitating longitudinal studies, enabling the objective tracking of disease progression or treatment response over time by consistently and accurately measuring changes in SC morphology [47]. Finally, and critically, accurate SC segmentation provides the necessary basis to co-register individual subject data to standardized templates, bringing data into a common coordinate space. This spatial normalization is essential for enabling group-level quantitative analysis, statistical comparisons across subjects, and the integration of data from different studies.

In a comprehensive review published by De Leener et al [48], automatic and semi-automatic SC segmentation methods were broadly categorized into distinct approaches. Intensity-based methods rely on the inherent signal intensity differences within the image to differentiate tissues. These techniques often involve feature extraction based on intensity characteristics, thresholding to separate voxels based on intensity ranges, and edge detection algorithms to identify boundaries based on intensity gradients [49]. Surface-based methods, in contrast, focus on explicitly modeling the boundaries of the SC as surfaces that evolve and adapt to the image data. Examples include active contours which are curves that deform to minimize an energy function, and deformable models, which extend this concept to 3D surfaces, allowing for more sophisticated shape representation and adaptation [50]. Image-based methods leverage prior knowledge and image similarity to guide segmentation. This includes template-based methods and clustering and classifier-based methods. Template-based methods are which utilize pre-segmented anatomical templates that are registered to the target image, transferring labels based on spatial correspondence Clustering and classifier-based methods employ machine learning algorithms to learn patterns and features from training data to classify voxels into different tissue types based on their image characteristics and spatial context.

2.3.3 Medical Image Segmentation in Spinal Cord Functional Magnetic Resonance Imaging

MIS in SC fMRI is crucial for achieving reproducible and robust group-level analyses. Spatial normalization, the process of aligning individual subject data to a common template space, is a prerequisite for voxel-wise statistical comparisons across a cohort, and accurate segmentation of the SC is a cornerstone of this normalization process. While segmentation can be performed on high-resolution structural images and then propagated to functional space via registration, performing segmentation directly within the functional space offers several advantages. Firstly, it ensures that the region of interest (ROI) definition is intrinsically aligned with the functional data itself, potentially minimizing misregistration errors that can accumulate when warping segmentations from structural to functional space. This direct functional space segmentation can enhance the accuracy of functional localization, ensuring that subsequent analyses are performed within the true boundaries of the SC as represented in the functional images. Second advantage is the reduced scan time. Getting an anatomical scan in the same session when the functional scan is acquired needs an extra overhead leading to possible discomfort of the subject and misalignment in between the two scans due to movement.

Currently, a common approach to SC fMRI analysis pipelines that are done using SCT, leverages a combination of anatomical and functional image segmentation. Typically, the pipeline initiates with the segmentation of the SC in high-resolution T2-weighted anatomical images using automated tools such as the *sct_deepseg_sc* [38] function in SCT. Manual labeling of vertebral levels, for example, C3 and C7, on these anatomical images is then performed to establish anatomical landmarks. These segmentations and vertebral labels are subsequently used to register the anatomical image to a standardized SC template like the PAM50 [51] template, utilizing functions such as *sct_register_to_template.* This anatomical registration yields a warping field defining the transformation to template space. For the functional data, often represented by temporal mean fMRI images, registration to the PAM50 template is performed using the *sct_register_multimodal* function. This step commonly incorporates manually contoured SC masks directly drawn on the temporal mean functional images. These manual masks serve as crucial regions of interest, guiding the multimodal registration and ensuring accurate alignment of the functional SC data with the template space, often initialized by the warping field derived from the anatomical registration. Voxels within these manual functional masks are typically weighted more heavily during the registration process, further refining the spatial normalization and prioritizing accurate alignment of the functional SC within the common template space. [17] use this mentioned pipleline in their work, especially to quantify the effects of inter-rater variability of manual masking in the registration process and highlight the need of improved image acquisition and postprocessing, standard preprocessing (segmentation methods for functional images). Other than pipeline, [45] were the only other who attempted to segment the SC directly on EPI data in a semiautomated fashion. Though they use k-means clustering for the segmentation, they have multiple preprocessing steps before the segmentation task.

2.3.4 Introduction to Convolutional Neural Networks

Convolutional Neural Networks (CNNs) [52] have become the dominant DL architecture for a wide range of computer vision tasks, including image classification, object detection, and, crucially for medical imaging, image segmentation [53]. Their architecture is inspired by the biological visual cortex and is designed to automatically and adaptively learn spatial hierarchies of features directly from input data, such as images [52]. This capability eliminates the need for manual, hand-crafted feature extraction, allowing CNNs to achieve state-of-the-art performance in complex image analysis problems [54].Within medical imaging, CNNs have driven significant progress in applications ranging from detecting pathologies in radiological scans to segmenting anatomical structures or lesions with high accuracy [55, 56].

The core computational unit of a CNN is the convolutional layer, which applies a set of learnable filters (kernels) across the input volume. Each filter detects specific local patterns (e.g., edges, corners, textures) at different spatial locations, generating feature maps that highlight the presence of these patterns [52]. Typically, these convolutional layers are followed by non-linear activation functions, such as the Rectified Linear Unit (ReLU), which introduces non-linearities essential for modeling complex data relationships [57]. Pooling layers, commonly max-pooling, are often interspersed within the network. These layers perform spatial downsampling, reducing the dimensionality of the feature maps, making the representations more robust to small translations and distortions, and decreasing computational load. By stacking multiple convolutional, activation, and pooling layers, CNNs build a hierarchy of features where initial layers capture simple patterns and deeper layers combine these to represent more complex and abstract concepts [52, 58]. For image classification, the final feature maps are typically flattened and fed into fully connected layers leading to an output layer (e.g., using a softmax function) that predicts class probabilities [54].

For medical image segmentation, the goal is pixel-wise classification. Architectures are adapted, often replacing the final fully connected layers with convolutional layers to preserve spatial information and produce an output segmentation map with the same spatial dimensions as the input [59]. This allows the network to assign a label (e.g., SC, tissue type, lesion presence) to each pixel (if 2D image) or voxel (if 3D image). Specialized architectures such as the U-Net, further refine this approach for biomedical segmentation by using an encoder-decoder structure with skip connections to effectively combine high-level contextual features with low-level spatial details leading to precise boundary delineation [60]. The capacity of CNNs to learn discriminative features hierarchically and perform dense predictions makes them exceptionally well-suited for the challenges of medical image segmentation.

2.3.5 Introduction to U-Nets

Building upon the foundational principles of Fully Convolutional Networks (FCNs) [59], the U-Net architecture was specifically developed to address the challenges inherent in biomedical image segmentation such as limited availability of annotated training data and the need for precise localization of object boundaries [60]. The U-Net quickly became a benchmark architecture in MIS due to its elegant design and remarkable effectiveness across various modalities and segmentation tasks [55, 61].

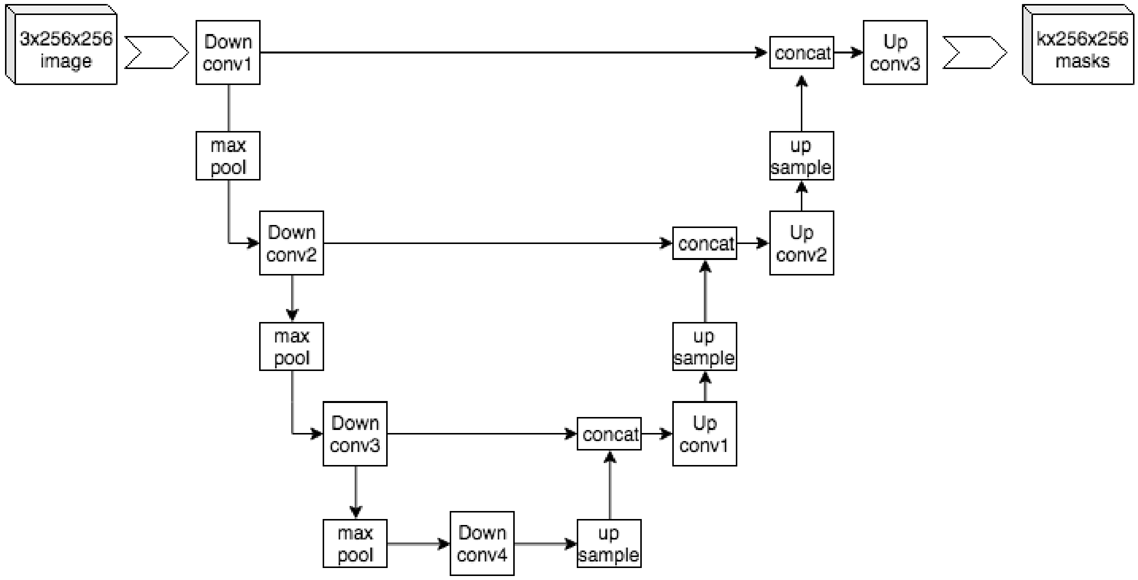

Figure 2.3 Architecture of UNet model. Source: Mehrdad Yazdani, Wikimedia Commons, CC Attribution-Share Alike 4.0 International license.

The distinctive U-shape (Fig. 2.3) of the network gives it its name and reflects its symmetric encoder-decoder structure. The encoder or contracting path follows the typical architecture of a CNN. It consists of repeated applications of two unpadded convolutions, each followed by a ReLU activation, and then a max pooling operation for downsampling. At each downsampling step, the number of feature channels is typically doubled. This path is responsible for capturing the context of the input image, learning increasingly complex features while progressively reducing spatial resolution [60].

The decoder or expanding path works symmetrically to the encoder. It consists of repeated modules, each starting with an up-convolution that halves the number of feature channels and doubles the spatial resolution. The resulting feature map is then concatenated with the correspondingly cropped feature map from the contracting path via skip connections. This concatenation is a key feature of the U-Net. Following the concatenation, standard convolutions each followed by ReLU are applied, similar to the contracting path. These skip connections allow the network to combine high-resolution spatial information from the early encoder layers (which capture fine details) with the abstract, contextual information learned in the deeper layers and propagated through the decoder [62]. At the final layer, a 1x1 convolution layer maps the feature vectors to the desired number of output classes.

A significant extension is the 3D U-Net, which adapts the architecture for volumetric segmentation by replacing 2D operations (convolutions, pooling) with their 3D counterparts, enabling direct processing of 3D medical images like MRI or CT volumes [63]. Due to its effectiveness and adaptability, the U-Net and its variants remain foundational models in numerous medical image segmentation applications, including brain tumor segmentation, organ delineation, cell tracking, and SC analysis [61].

To evaluate the segmentation performance, common metrics quantify the spatial agreement between the predicted segmentation (P) and the ground truth segmentation (G). The Dice Score (DS), perhaps the most widely used metric for volumetric overlap, measures the similarity between the two sets of voxels [64]. It is defined as:

$$\mathrm{DS}(P, G) = \frac{2|P \cap G|}{|P| + |G|}$$

where $|P \cap G|$ represents the number of voxels common to both the prediction and the ground truth, and $|P|$ and $|G|$ represent the total number of voxels in the prediction and ground truth, respectively. DS ranges from 0 (no overlap) to 1 (perfect overlap). Another important metric which is also relevent to this thesis is the Hausdorff Distance (HD) [65] which is particularly sensitive to boundary accuracy. It measures the maximum distance from a point in one set to the closest point in the other set. More formally, the directed HD from P to G is $h(P, G) = max_{p \in P}\{min_{g \in G}\{d(p, g)\}\}$, where $d(p, g)$ is the distance between points p and g. The bidirectional HD is then defined as:

$$\mathrm{HD}(P, G) = \max(h(P, G), h(G, P))$$

A lower HD value indicates a better match between the predicted and ground truth boundaries. Often, the 95th percentile of the distances (HD95) is reported to reduce sensitivity to outliers.

2.3.6 Human-in-the-Loop Active Learning

While DL models such as U-Nets have shown remarkable success in medical image segmentation, their performance heavily relies on the availability of large amounts of accurately annotated training data [55, 56]. In many medical domains, acquiring such largescale and highquality annotations is a significant bottleneck, as it typically requires substantial time and expertise from clinicians or trained annotators [66]. HIL-AL offers a compelling paradigm to mitigate this challenge by integrating the human expert directly into the model training loop [67, 68]. Instead of passively training on a randomly sampled or fully labeled dataset, HIL-AL focuses on iteratively selecting the most informative unlabeled data points and presenting them to a human expert (the "oracle") for annotation. This optimizes the use of limited expert time and maximizing model performance with fewer labeled samples [69].

The HIL-AL process functions as an iterative cycle, tightly coupling model training with expert feedback:

1. Initial Training: A model (e.g., a U-Net) is trained on a small initial set of expertly labeled data.

2. Querying: The current model analyzes a pool of unlabeled data. A query strategy entails the selection of a subset of samples the deem to be the most beneficial for the model to learn from next. Strategies like manual selection by human-expert, uncertainty sampling [70] or query-by-committee [71] are employed to select the next batch of samples.

3. Human Annotation: The selected samples are presented to the human expert. The expert provides accurate labels or corrects the model's current predictions for these specific samples.

4. Iteration: The cycle (steps 2-4) repeats, progressively refining the model's accuracy by focusing expert effort on the most impactful data points until a desired performance level is met [67].

2.3.7 Catastrophic Forgetting

A significant challenge arises when training these models such as CNNs. U-Nets, in a sequential manner, following the HIL-AI strategy. While these networks effectively learn when trained concurrently, with all training items repeatedly presented in a mixed fashion, they often exhibit a phenomenon known as catastrophic forgetting when trained sequentially on distinct tasks or datasets. This term, prominently highlighted by McCloskey and Cohen [72], refers to the tendency of a network to abruptly and drastically lose the ability to perform previously learned tasks upon learning a new one.

This problem is particularly pronounced in networks that utilize distributed representations, where information about multiple items or concepts is encoded across the same set of connection weights. When the network adjusts these shared weights to learn a new task (Task B), the modifications overwrite or interfere with the weight configurations necessary to perform an earlier task (Task A), leading to a rapid degradation or complete loss of the previously acquired knowledge [72, 73].

Addressing this limitation is crucial for developing AI systems that can adapt and acquire new knowledge without detrimentally impacting prior learning, a capability essential for realworld applications. In our case, since we train the model in a sequential fashion, we observe a drop in performance after a few iterations, specifically when we start focusing on low quality data.

CHAPTER 3 OVERALL METHODOLOGY

The article presented in this thesis is "EPISeg: Automated segmentation of the SC on echo planar images using open-access multi-center data". This article has been submitted to the the Imaging NeuroScience journal in January 2025. An abstract version of this article has already been accepted at International Society for Magnetic Resonance in Medicine (ISMRM) 2025 Annual Meeting in Februrary 2025 and has been invited for an oral presentation (top 10%) in May 2025.

The core contribution of this work, detailed in the Chapter 4 is the development and validation of EPISeg, a novel DL-based method designed specifically for this challenging task. This involved two major interconnected contributions:

- Creation of a Large-Scale Open-Access Dataset: To enable the development and robust validation of EPISeg, we curated and publicly released a unique, large-scale (N=406 participants) multi-center (15 sites) dataset of SC gradient-echo EPI images. This dataset, organized according to the Brain Imaging Data Structure (BIDS) standard, includes associated expert-verified ground-truth segmentations. Made available via OpenNeuro, this dataset represents a valuable resource for the community, facilitating future research and development in SC fMRI analysis and segmentation.
- Development of EPISeg: We propose and validate EPISeg, an automated DL model (based on the 3D U-Net architecture) for segmenting the SC directly on EPI data. Addressing the scarcity of labeled data, the model was trained using an iterative HIL-AL strategy, enabling efficient refinement and enhanced robustness against diverse image artifacts and acquisition protocols. The final EPISeg model demonstrates improved segmentation accuracy compared to existing methods and has been integrated into the Spinal Cord Toolbox (SCT) for accessibility by the research community.

CHAPTER 4 ARTICLE 1: EPISEG: AUTOMATED SEGMENTATION OF THE SPINAL CORD ON ECHO PLANAR IMAGES USING OPEN-ACCESS MULTI-CENTER DATA

4.1 Article metadata

- **Title:** EPISeg: Automated segmentation of the spinal cord on echo planar images using open-access multi-center data
- **Authors:** Rohan Banerjee[1,2,27], Merve Kaptan[3], Alexandra Tinnermann[4], Ali Khatibi[5,6,7,8], Alice Dabbagh[10], Christian Büchel[4], Christian W. Kündig[11], Christine S.W. Law[3], Dario Pfyffer[3,11], David J. Lythgoe[16], Dimitra Tsivaka[16,34], Dimitri Van De Ville[19,20], Falk Eippert[10], Fauziyya Muhammad[21], Gary H. Glover[14], Gergely David[11], Grace Haynes[33], Jan Haaker[4], Jonathan C. W. Brooks[18], Jürgen Finsterbusch[4], Katherine T. Martucci[15], Kimberly J. Hemmerling[22,23], Mahdi Mobarak-Abadi[6], Mark A. Hoggarth[22,24], Matthew A. Howard[16], Molly G. Bright[22,23], Nawal Kinany[19,20], Olivia S. Kowalczyk[13,16], Patrick Freund[11,12,13], Robert L. Barry[25,26,30], Sean Mackey[3], Shahabeddin Vahdat[5,6,31], Simon Schading[11], Stephen B. McMahon[17], Todd Parish[23,32], Véronique Marchand-Pauvert[9], Yufen Chen[23], Zachary A. Smith[21], Kenneth A. Weber II[3], Benjamin De Leener[1,27,29†], Julien Cohen-Adad[2,27,28,29*†]
- **Affiliations:** [1]Department of Computer Science, Polytechnique Montreal, Montreal, Quebec, Canada [2]Mila - Quebec AI Institute, Montreal, Quebec, Canada [3]Division of Pain Medicine, Department of Anesthesiology, Perioperative and Pain Medicine, Stanford University School of Medicine, Palo Alto, CA, United States, [4]Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Germany, [5]McConnell Brain Imaging Center, Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada, [6]Centre de recherche de l'Institut Universitaire de Gériatrie de Montréal, Montréal, Quebec, Canada, [7]Centre of Precision Rehabilitation for Spinal Pain (CPR Spine), University of Birmingham, UK, [8]Centre for Human Brain Health, University of Birmingham, UK, [9]Sorbonne Université, Inserm, CNRS, Laboratoire d'Imagerie Biomédicale, LIB, F-75013, Paris, France, [10]Max Planck Research Group Pain Perception, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany, [11]Spinal Cord Injury Center, Balgrist University Hospital, University of Zurich, Zurich, Switzerland, [12]Department of Neurophysics, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany, [13]Functional

Imaging Laboratory, Department of Imaging Neuroscience, Queen Square Institute of Neurology, University College London, London, UK, [14]Radiological Sciences Laboratory, Department of Radiology, Stanford University School of Medicine, Palo Alto, CA, United States, [15]Center for Translational Pain Medicine, Department of Anesthesiology, Duke University Medical Center, Durham, NC, USA, [16]Department of Neuroimaging, Institute of Psychology, Psychiatry & Neuroscience, King's College London, London, UK, [17]Wolfson Centre for Age Related Diseases, King's College London, London UK, [18]Department of Psychology, University of Liverpool, Liverpool, United Kingdom, [19]Department of Radiology and Medical Informatics, University of Geneva, Geneva, Switzerland, [20]Neuro-X Institute, Ecole Polytechnique Fédérale de Lausanne (EPFL), Geneva, Switzerland, [21]Department of Neurosurgery, University of Oklahoma Health Sciences Center, OK, USA, [22]Department of Physical Therapy and Human Movement Sciences, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA, [23]Department of Biomedical Engineering, McCormick School of Engineering, Northwestern University, Evanston, Illinois, USA, [24]Department of Physical Therapy, North Central College, Naperville, Illinois, USA, [25]Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Charlestown, MA, USA, [26]Harvard Medical School, Boston, MA, USA, [27]NeuroPoly Lab, Institute of Biomedical Engineering, Polytechnique Montreal, Montreal, Quebec, Canada, [28]Functional Neuroimaging Unit, CRIUGM, University of Montreal, Montreal, Quebec, Canada, [29]Research Center, Ste-Justine Hospital University Centre, Montreal, Quebec, Canada, [30]Harvard-Massachusetts Institute of Technology Health Sciences & Technology, Cambridge, MA, USA, [31]Department of Applied Physiology and Kinesiology, University of Florida, Gainesville, FL, USA, [32]Department of Radiology, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA, [33]Stephenson School of Biomedical Engineering at the University of Oklahoma in Norman, OK, USA, [34]Medical Physics Department, Medical School, University of Thessaly, Larisa, Greece

- **Submitted to:** Imaging Neuroscience on 27 January 2025
- **Accepted at:** ISMRM 2025; Oral presentation (top 10%)
- **Preprint DOI:** `https://doi.org/10.1101/2025.01.07.631402`
- **Dataset url:** `https://openneuro.org/datasets/ds005143`

- **Contributions by Rohan Banerjee**: Rohan Banerjee contributed nearly 85% towards the development of the presented article. These contributions included - Curation and releasing the open-source dataset, conceptualizing the methodology, carrying out formal analyses, development of open-source code, validating the methodology, writing the original draft along with incorporating co-author suggestions and creating tables and figures for the article.

4.2 Abstract

Functional magnetic resonance imaging (fMRI) of the spinal cord is relevant for studying sensation, movement, and autonomic function. Preprocessing of spinal cord fMRI data involves segmentation of the spinal cord on gradient-echo echo planar imaging (EPI) images. Current automated segmentation methods do not work well on these data, due to the low spatial resolution, susceptibility artifacts causing distortions and signal drop-out, ghosting, and motion-related artifacts. Consequently, this segmentation task demands a considerable amount of manual effort which takes time and is prone to user bias. In this work, we (i) gathered a multi-center dataset of spinal cord gradient-echo EPI with ground-truth segmentations and shared it on OpenNeuro `https://openneuro.org/datasets/ds005143/versions/1.3.0`, and (ii) developed a deep learning-based model, EPISeg, for the automatic segmentation of the spinal cord on gradient-echo EPI data. We observe a significant improvement in terms of segmentation quality compared to other available spinal cord segmentation models. Our model is resilient to different acquisition protocols as well as commonly observed artifacts in fMRI data. The training code is available at `https://github.com/sct-pipeline/fmri-segmentation/`, and the model has been integrated into the Spinal Cord Toolbox as a command-line tool.

4.3 Keywords

Spinal Cord, Echo Planar Imaging, Functional Magnetic Resonance Imaging, Segmentation

4.4 Introduction

Functional magnetic resonance imaging (fMRI) allows non-invasive investigation of the central nervous system, of which, the spinal cord (SC) is a crucial part [8], [10], [74]. SC fMRI has the potential for tracking disease progression and assessing treatment effectiveness by providing novel insights into various neurological conditions such as multiple sclerosis, neuropathic pain, and SC injury [75], [76], [77], [78]. Although fMRI was introduced in the 1990s, the application of fMRI in the SC has been lagging behind brain imaging studies due to the technical challenges owing to data acquisition and processing [79], [14]. Despite these difficulties and thanks to recent technical developments, SC fMRI is rapidly taking off, with applications to study sensorimotor systems, as well as autonomic processes in humans and animals [5], [7], [80], [81], [82], [83], [84], [31], [85], [3], [86], [87], [88].

FMRI data are commonly acquired with a gradient-echo echo planar imaging (EPI) sequence, due to its short acquisition times (about 2s per volume) and its heavy T2* weighting, which produces increased sensitivity to the blood oxygenation level contrast [89]. The most notable challenge arises from the substantial changes in the magnetic susceptibility of tissues surrounding the SC, which introduces image distortions and signal drop-outs in the images [10], [8], [79], [14]. Additionally, physiological noise, motion artifacts, and poor contrast between the SC and surrounding pulsatile cerebrospinal fluid lead to degradation in the quality of EPI images [90], [15].

The analysis of fMRI data typically involves statistical inferences at the group level, which necessitates the valid and precise normalization of data into a template space which aims to establish a one-to-one correspondence between the data from different individuals. The emergence of SC templates [51], [91], [92], [93], [94] and normalization procedures [32], [95], [92] make the group-level analysis accessible and reproducible across different SC fMRI datasets.

Currently, the gold standard for registering SC data to a template relies on the segmentation of the SC [46]. While there has been important development of automatic methods to segment the SC on MRI images [37], [36], [38], [35], [39], [40], [41], [42], [43], [44], these methods do not work well on gradient-echo EPI data due to aforementioned challenges. Moreover, the relatively low resolution of axial EPI scans (typically 1-1.5 mm in-plane, with 3-5mm thickness) makes the segmentation task even harder. An automated and robust segmentation method for the SC is therefore desirable to increase the reproducibility of SC fMRI studies [17].

In this work, we introduce EPISeg, a deep learning (DL) based automatic SC segmentation method. Along with EPISeg, this paper also introduces open-access fMRI data of the SC from multiple sites (N=15). The dataset comes with ground-truth SC segmentations that were used to train the EPISeg model. We demonstrate the robustness of EPISeg amongst different sites, involving different scanners including data from healthy volunteers as well as patient groups, and against commonly observed artifacts such as partial volume effects, and ghosting. We compare our method with other state-of-the-art segmentation models like PropSeg [36], DeepSeg [38] and Contrast-agnostic SC segmentation [35], and show a significant improvement in segmentation performance using the Dice Score (DS) and Hausdorff Distance (HD). To our best knowledge, this is the first work regarding automatic segmentation of SC on fMRI data. The contributions of this paper are:

1. Creation of a new open-source SC gradient-echo EPI dataset (n=406 participants, 15 sites).
2. Introduction of an automated segmentation method for SC on fMRI data, thoroughly validated on diverse datasets spanning different populations and different regions of the SC including the lumbar and cervical SC.
3. Integration of this algorithm as a single-line command in SCT [32], making it easily accessible to the scientific community.

4.5 Methods

In this section, we describe the data collection, manual labeling, model architecture, and model training using an active learning framework. An overview of the pipeline is shown in Figure 4.1.

4.5.1 Data

We gathered gradient-echo EPI datasets from multiple sites (15 sites, 406 participants in total). All volunteers provided written informed consent following Institutional Review Board approval and the Declaration of Helsinki (see Table 4.1 for details of acquisitions for all sites). All participating sites were asked to share (i) the first 20 motion-corrected EPI volumes of the functional run (more volumes were unnecessary as the primary goal of these images is to perform segmentation, not fMRI analysis), (ii) the BIDS-compatible JSON file containing information about the acquisition, (iii) the mean image of the motion-corrected data, and (iv) the SC segmentation based on the mean motion-corrected image. Some sites manually created the segmentations from scratch, while other sites started off from a segmentation generated by automated methods such as PropSeg or DeepSeg, with manual correction. Out of 406 volunteers across all sites, the majority were healthy volunteers (n=359), 17 volunteers had degenerative cervical myelopathy, and 30 volunteers had fibromyalgia.

The dataset was put together into one multi-site dataset and organized according to the BIDS convention [33]. The segmentations produced at each site were reviewed by three raters overseeing the entire study (RB, MK, JCA) using the sct_qc quality control (QC) tool [46]. Segmentations were subsequently updated using EPISeg wherever we found erroneous segmentations. The dataset was uploaded on OpenNeuro `https://openneuro.org/dataset/ds005143/versions/1.3.0`, except for one site due to the data sharing policy of their respective ethics committee.

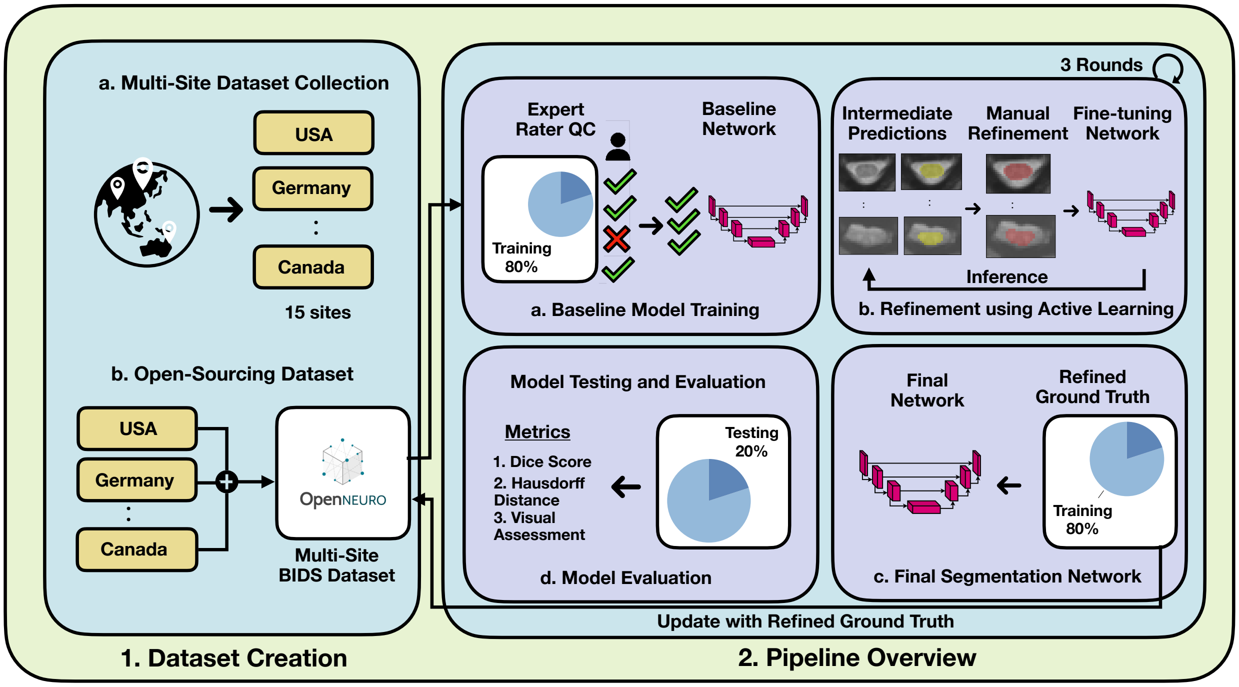

Figure 4.1 **EPISeg overall pipeline.** 1. Dataset creation and open-sourcing process. 1a. Initiation of dataset contribution call for EPISeg. 1b. The collected datasets are rearranged according to the brain imaging data structure (BIDS) convention and uploaded to OpenNeuro, forming a multi-site BIDS dataset. 2. Training and validation pipeline overview. 2.a. An expert selects reliable segmentations, which are used to train the baseline model. This model undergoes expert rater quality control (QC) to ensure high-quality training data. 2b. The baseline model generates initial predictions, which are then manually refined by experts. These refined predictions are used to fine-tune the baseline model. This process is repeated for three rounds, ensuring incremental improvement and robustness. 2c. Participants and refined ground truth from the refinement phase are used to train the final network. The refined ground truth is also updated in the open-source dataset. 2d. The final network is used for inference on test data, with metrics such as Dice Score (DS), Hausdorff Distance (HD), and visual assessments reported for comprehensive evaluation.

*HV = Healthy Volunteers; FoV = Field of View; TR = Repetition Time; TE = Echo Time; FA = Flip Angle; ★ = Semi-automatic segmentation; ⧫ = Automatic segmentation

Table 4.1 Dataset details

Site name	Institution	Sample	Scanner and sequence
Leipzig_rest★ [96, 97]	Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany	HV; N=45 (20F); Age: 27±3.8	3T whole-body Siemens Prisma MRI System; 2D gradient-echo EPI; Number of slices: 24; in-plane resolution: $1.0 \times 1.0\text{mm}^2$; slice thickness: 5.0mm; FoV = $128 \times 128\text{mm}^2$; TR: 2.312s; TE: 40 ms; FA = 84°
Geneva_rest1 [4]	École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland	HV; N=17 (11F); Age: 27±3.5	3T whole-body Siemens Prisma MRI System; 2D gradient-echo EPI; number of slices: 27; FoV = $48 \times 144\text{mm}^2$; in-plane resolution = $1 \times 1\text{mm}^2$; slice thickness = 3mm; TR = 2.5s; TE = 34 ms; FA = 80°
Geneva_rest2 [7]	École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland	HV; N=12 (6F); Age: 28.8±3.2	3T whole-body Siemens Prisma MRI System; 2D gradient-echo EPI; number of slices: 27; FoV = $48 \times 144\text{mm}^2$; in-plane resolution = $1 \times 1\text{mm}^2$; slice thickness = 3mm; TR = 2.5s; TE = 34 ms; FA = 80°
Geneva_rest3 [8]	École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland	HV; N=14 (7F); Age: 28.6±4.2	3T whole-body Siemens Prisma MRI System; 2D gradient-echo EPI; number of slices: 31; FoV = $48 \times 144\text{mm}^2$; in-plane resolution = $1 \times 1\text{mm}^2$; slice thickness = 3mm; TR = 2.5s; TE = 34 ms; FA = 80°

Continued on next page

Table 4.1 Dataset details (continued)

Site name	Institution	Sample	Scanner and sequence
KCL_rest★ [98]	King's College London, London, UK	HV; N=23 (13F); Age: 23.9±3.8	3T General Electric MR750 System; 2D gradient-echo EPI; number of slices: 38; FoV = 180mm; reconstruction matrix = 128 × 128mm^2; in-plane resolution = 1.875 × 1.875mm^2; slice thickness = 4mm (1mm gap); TR = 2.5s; TE = 30 ms; FA = 90°
NW_motor [84]	Northwestern University, Chicago, Illinois, USA	HV; N=28 (18F); Age: 26±4.7	3T whole-body Siemens Prisma MRI System; 2D gradient-echo EPI; number of slices: 25; FoV = 128 × 44mm^2; in-plane resolution = 1 × 1mm^2; slice thickness = 3mm; TR = 2s; TE = 30ms; FA = 90°
NW_motor-Weber [3]	Northwestern University, Chicago, Illinois, USA	HV; N=11 (6F); Age: 27.7±1.9	3T whole-body Siemens Prisma MRI System; 2D gradient-echo EPI; number of slices: 31; FoV = 128 × 44mm^2; in-plane resolution = 1 × 1mm^2, slice thickness = 3mm, TR = 2.5s, TE = 30ms, FA = 80°
NW_thermal [86]	Northwestern University, Chicago, Illinois, USA	HV; N=12 (3F); Age: 28.8±2.5	3T whole-body Siemens Prisma MRI System; 2D gradient-echo EPI; number of slices: 31; FoV = 128 × 128mm^2; in-plane resolution = 1 × 1mm^2, slice thickness = 3mm, TR = 2.5s, TE = 30ms, FA = 80°

Continued on next page

Table 4.1 Dataset details (continued)

Site name	**Institution**	**Sample**	**Scanner and sequence**
NW_tactile [82]	Northwestern University, Chicago, Illinois, USA	HV; N=12 (3F); Age: 28.8±2.5 Degenerative cervical myelopathy N = 2; Age: 33.5 ± 2.1	3T whole-body Siemens Prisma MRI System; 2D gradient-echo EPI; number of slices: 25; FoV = 128 × $44mm^2$; in-plane resolution = 1 × $1mm^2$; slice thickness = 3mm; TR = 2s; TE = 30ms; FA = 80°
Zurich_Cerv_Bilatmotor	Balgrist University Hospital, University of Zurich, Zurich, Switzerland	HV; N=18 (3F); Age: 28.8±2.5	3T whole-body Siemens Prisma MRI System; 2D gradient-echo EPI; number of slices: 18; FoV = 128 × $128mm^2$; in-plane resolution = 1 × $1mm^2$; slice thickness = 5mm; TR = 3.26s; TE = 31ms; FA = 87°
Stanford_rest [99]	Stanford University, Palo Alto, California, USA	HV; N=29 (F); Age: 25.6±5.5	GE 3T Discovery 750; 2D gradient-echo EPI; number of slices: 43; FoV = 80mm; matrix = 64 × $64mm^2$; in-plane resolution = 1.25 × $1.25mm^2$; slice thickness = 5mm; TR = 2.5s; TE = 30ms; FA = 80°
Stanford_restmartucci [100, 101]	Stanford University, Palo Alto, California, USA	HV; N=15 (F); Age: 49±10.8 Fibromyalgia N=30 (F); Age: 50.2±8.8	3T General Electric MR750 System; number of slices: 14; FoV = 160mm; reconstruction matrix = 128 × $128mm^2$; in-plane resolution = 1.25 × $1.25mm^2$; slice thickness = $4mm^2$; TR = 2.5s; TE = 30ms; FA = 75°

Continued on next page

Table 4.1 Dataset details (continued)

Site name	**Institution**	**Sample**	**Scanner and sequence**
UNF_MSL [85, 102]	Centre de recherche de l'Institut Universitaire de Gériatrie de Montréal, Montréal, QC, Canada	HV; N=27 (14F); Age: 24.9±3	3T whole-body Siemens TIM Trio MRI System; 2D gradient-echo; number of slices: 40-43 (8-10 SC); FoV = $132 \times 132\text{mm}^2$; in-plane resolution = $1.2 \times 1.2\text{mm}^2$; slice thickness = 5mm (9mm gap); TR = 3.14s; TE = 33ms; FA = 90°
OUHSC_stim	University of Oklahoma Health Sciences Center, OK, USA	HV; N=13(9F); Age: 50.3±6.6 Cervical Spondylotic Myelopathy N=15(10F); Age: 57±9.9	3T whole-body Siemens Prisma MRI System; 2D gradient-echo EPI; number of slices: 10; FoV = $128 \times 128\text{mm}^2$; in-plane resolution = $1 \times 1\text{mm}^2$; slice thickness = 5mm; TR = 0.86s; TE = 30ms; FA = 50°
Hamburg_pain⧫ [103]	Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Germany	HV; N=30 (M); Age: 23.9±3.8	3T whole-body Siemens Prisma MRI System; 2D gradient-echo EPI; number of slices: 10; FoV = $128 \times 128\text{mm}^2$; in-plane resolution = $1 \times 1\text{mm}^2$; slice thickness = 5mm; TR = 0.86s; TE = 30ms; FA = 50°
Zurich_Lumb_rest [88]	Balgrist University Hospital, University of Zurich, Zurich, Switzerland	HV; N=13 (4F); Age: 28.7±3.5	3T whole-body Siemens Prisma MRI System; 2D gradient-echo EPI; number of slices: 15; FoV = $144 \times 48\text{mm}^2$; in-plane resolution = $1 \times 1\text{mm}^2$; slice thickness = 5mm; TR = 1.29s; TE = 35ms; FA = 74°

Continued on next page

Table 4.1 Dataset details (continued)

Site name	**Institution**	**Sample**	**Scanner and sequence**
Leipzig_pain♦ [104]	Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany	HV; N=40 (20F); Age: 27.3±4.6	3T whole-body Siemens Prisma MRI System; 2D gradient-echo EPI; number of slices: 16; in-plane resolution:$1.0 \times 1.0\text{mm}^2$; slice thickness: 5.0mm; FoV: $128 \times 128\text{mm}^2$; TR: 1.8s; TE: 40ms; FA = 70°

4.5.2 Data preprocessing

Our proposed model is designed to segment the SC on fMRI images. The input consists of a mean motion-corrected volume representing the fMRI scan. This mean volume condenses the motion-corrected time-series data into a single three-dimensional scan by averaging over all time points. Each input volume is matched with a ground truth binary segmentation that outlines the spinal cord's boundaries.

4.5.3 Criteria for selecting training data

When performing multi-step active learning training, the first training batch necessitates reliable ground truth segmentation. In this case, we aimed to identify data subsets, D_g, that had accurate ground truth SC segmentations. These data subsets were in general of high quality, notably thanks to the z-shimming method [105] used in some of the datasets like the Leipzig_pain, Leipzig_rest. It is important to note that all selection processes prioritized accurate segmentations over image quality (which is inherently related to segmentation quality), ensuring that the model was trained on the most precise delineations of the SC.

4.5.4 Model training protocol

Our training protocol involves a three-step model training procedure employing a human-in-the-loop active learning to iteratively remove errors and improve ground truth accuracy and then training a final segmentation model which is deployed in SCT. Our reasoning for using active learning in this setting was to develop a model resilient to the various types of image artifacts it encounters over the different rounds of training.

In the first step, as shown in Figure 4.2 we trained a baseline model M_b using the dataset D_g which had 96 participants. The baseline model is a 3D U-Net [63], trained using the

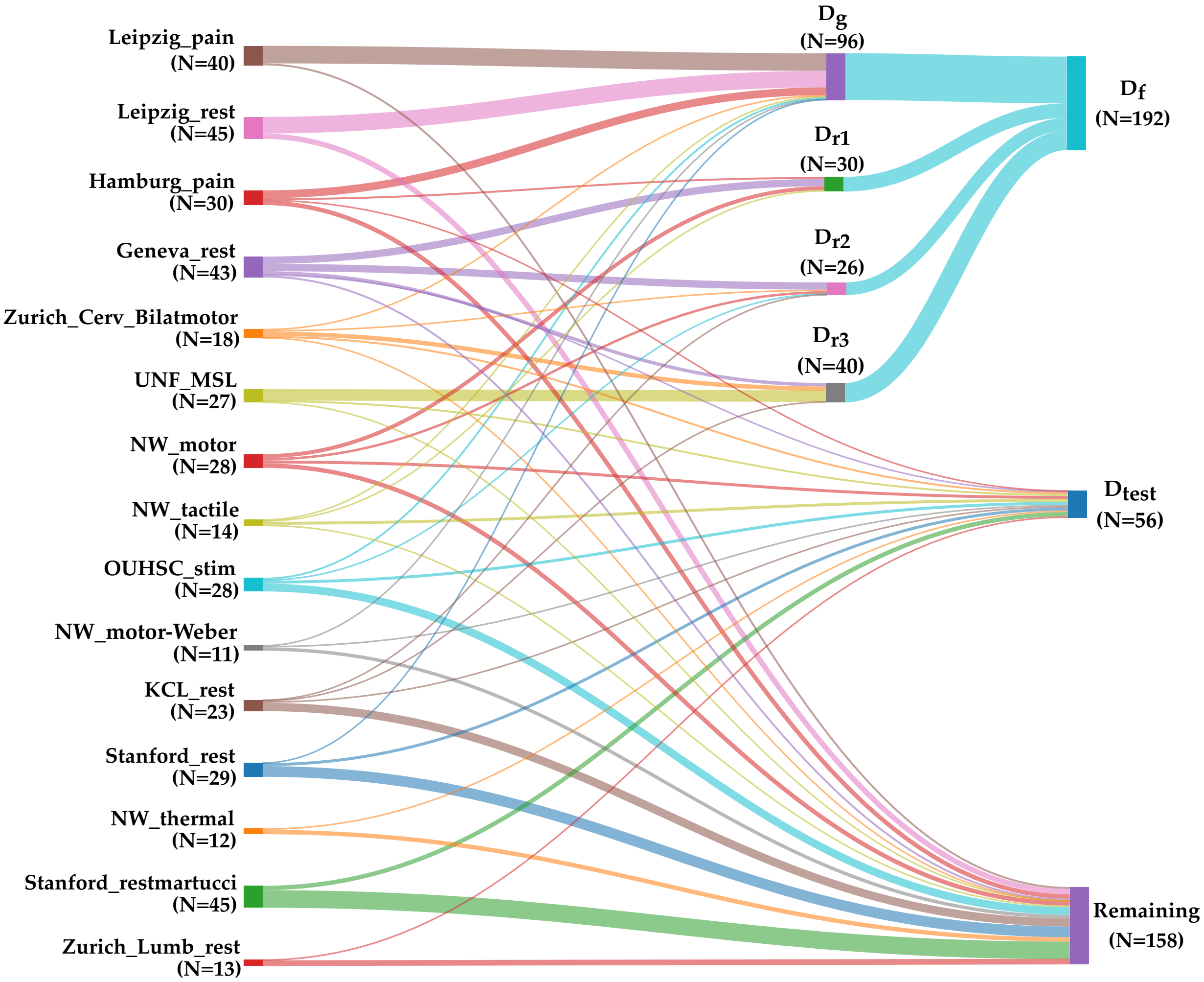

Figure 4.2 **Dataset split.** The sankey chart shows the different dataset splits created during the training process. The split D_g is for the baseline model, D_{r1} is for the first round of active learning training, D_{r2} is for the second round of active learning training, D_{r3} is for the third round of active learning training, D_f is the aggregated set of D_g, D_{r1}, D_{r2}. D_{r3} and D_{test} is the held-out test set to validate the EPISeg performance. The "remaining" set contains images with artifacts, erroneous segmentations.

nnUNetv2 framework [106], which self-configures parameters for optimal performance. We chose nnUNetv2 to perform this task because it has shown state-of-the-art performance in a range of medical anatomies and imaging modalities [106], [107]. For training M_b, we used the default parameters configured by nnUNetv2 for our dataset. The training details include a batch size of 5, learning rate of 0.01 with a polynomial learning rate scheduler and a stochastic gradient descent optimizer for a total of 1,000 epochs. The model was trained using the combination of Dice loss and cross-entropy loss.

The second step consisted of iterative refinement of training ground truth and model using active learning. Once the baseline model, M_b was trained, we used it to generate segmentations on the remaining dataset. For the first round of active learning, a subset of these segmentations, D_{r1}, consisting of 30 images were chosen for quality control assessment and manual corrections, if necessary. This subset D_{r1} was then used to fine-tune M_b, resulting in an updated model in round 1 of training, M_{r1}. To further improve the model's capability, two more rounds of active learning with the same procedure with variable dataset sizes were conducted which resulted in the creation of datasets D_{r2}, D_{r3} and corresponding models M_{r2} and M_{r3} respectively. These training rounds used the same default parameters used to train M_b.

For the third and final step of the training protocol, all the datasets D_g, D_{r1}, D_{r2} and D_{r3} were aggregated into one final dataset, D_f which had 192 3D images to train one model, M_f. Our reasoning behind training the M_f was to overcome the signs of forgetting that M_{r3} was exhibiting. Forgetting is a common phenomenon that is observed during active learning i.e. when networks are trained in a sequential fashion, the new learnings of the network tend to interfere with the old learning of the network [72]. We discuss further about this observation in the "Effect of active learning on EPISeg" results section. Following a similar training procedure from the first step, a new set of nnUNetv2 parameters were configured for D_f, which included a batch size of 15 with a polynomial learning rate scheduler, trained for 1,000 epochs. Along with this, we use the latest nnUNetv2 ResEnc [107] variant of the UNet with residual connections in the encoder since the authors mention that using residual connections along with the nnUNetv2 framework is a recipe for better segmentation performance.

4.5.5 Evaluation protocol

To evaluate our model performance, we created a multi-site held-out test set, D_{test}. The whole dataset (containing all subjects from all sites) was split in a 80:20 ratio where participants from the 80% of the dataset were used for training and 20% of the dataset were held out for testing. The D_{test} was not included in the training or validation sets. To have a fair representation of all the sites in D_{test}, we randomly chose 20% of the participants from each site and aggregated it into one dataset which in total had 56 participants. We evaluate our model's performance on the D_{test} and compare it with the SC segmentation models PropSeg, DeepSeg and Contrast-agnostic SC segmentation model. We discuss these results in the following sections.

We use Dice Score (DS) and Hausdorff Distance (HD) as validation metrics. For HD, we use the 95th percentile HD (HD95) since it reduces the impact of outlier errors, providing a more robust and representative measure of segmentation accuracy. The DS quantifies the spatial agreement between the predicted and ground truth segmentations, with higher scores indicating a closer alignment. On the other hand, the HD measures the maximum separation between the predicted SC segmentation and the nearest point in the ground truth segmentation, providing insights into the model's ability to capture the geometric correspondence between predicted and true SC boundaries. We utilize the implementations provided by the MetricsReloaded framework [108] for our metrics. Additionally, we verified the model's performance to ensure that model was producing segmentations as expected [discussed in section 4.6.4]

4.6 Results

4.6.1 Automatic fMRI SC segmentation

Table 4.2 compares the performance of the EPISeg models trained on various datasets - M_{r3} trained on the final round of the active learning data subset D_{r3}; M_f trained on the aggregated set D_f from all rounds of active learning rounds; M_{ResEnc} is the model with ResNet connections trained on D_f. The best performance (in terms of DS) is seen with the M_f dataset, indicating the advantage of having a single model with reliable segmentations rather than having several rounds of fine-tuning.

Table 4.2 Performance of the proposed EPISeg on D_{test}. M_{r3} refers to the model obtained after the 3^{rd} round of active learning training, M_f refers to the model trained on D_f, and M_{ResEnc} refers to the latest ResNet-based nnUNet model trained on D_f.

Model	**DS ± S.D.**	**HD95 ± S.D. (in mm)**
EPISeg (M_{r3})	0.82 ± 0.06	1.73 ± 1.03
EPISeg (M_f)	0.88 ± 0.05	1.41 ± 0.86
EPISeg (M_{ResEnc})	0.87 ± 0.05	1.28 ± 0.73

Even though we see M_f performs a little better than M_{ResEnc} in terms of DS, M_{ResEnc} has a better HD95 score. When we investigated this qualitatively, we saw that M_{ResEnc} gave more complete segmentations, for example, it was able to segment the first and last slice more consistently and was more robust to slices with artifacts. We therefore used the M_{ResEnc} as our final EPISeg model.

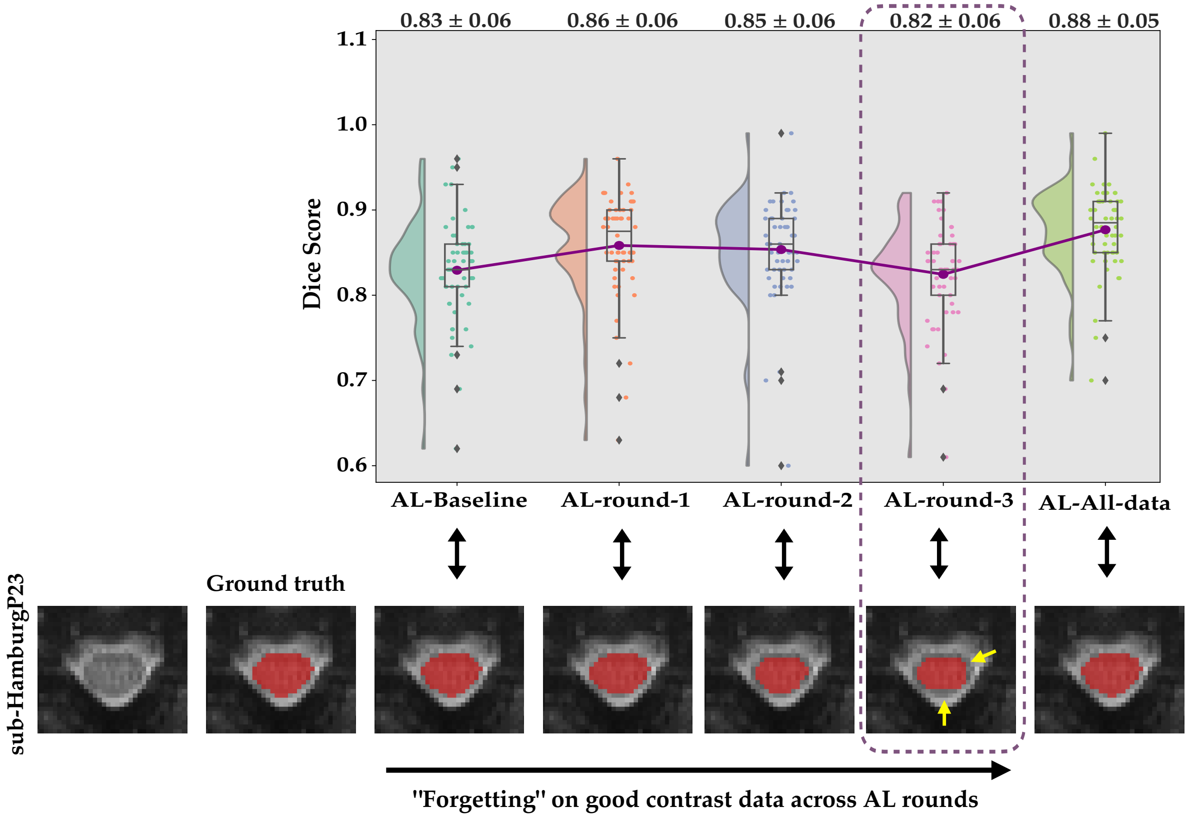

Figure 4.3 **Effects of Active Learning.** The raincloud plot shows the DS distribution of different models trained during the active learning training rounds. The images below show the axial slice from a subject and the red overlays show the model prediction of the SC. The dotted line and the yellow arrows show the performance degradation of the model over the training rounds.

4.6.2 Effect of active learning on EPISeg

We relied heavily on human-in-the-loop active learning for training EPISeg. Our selection of D_g was based on the quality of the segmentations rather than the image quality. However, since it was easier to achieve better segmentations on higher quality images, they were predominantly used in training M_b. More and more lower quality images were introduced in the later rounds of active learning. As we can observe in Figure 4.3, there is a noticeable dip in the performance of M_{r3}. When we investigated this phenomenon, we observed that, while the model's performance improved for lower quality images, it degraded the performance for high quality images similar to the ones used in training M_b. We refer to this as forgetting and therefore decided to train M_f to overcome this drawback of EPISeg.

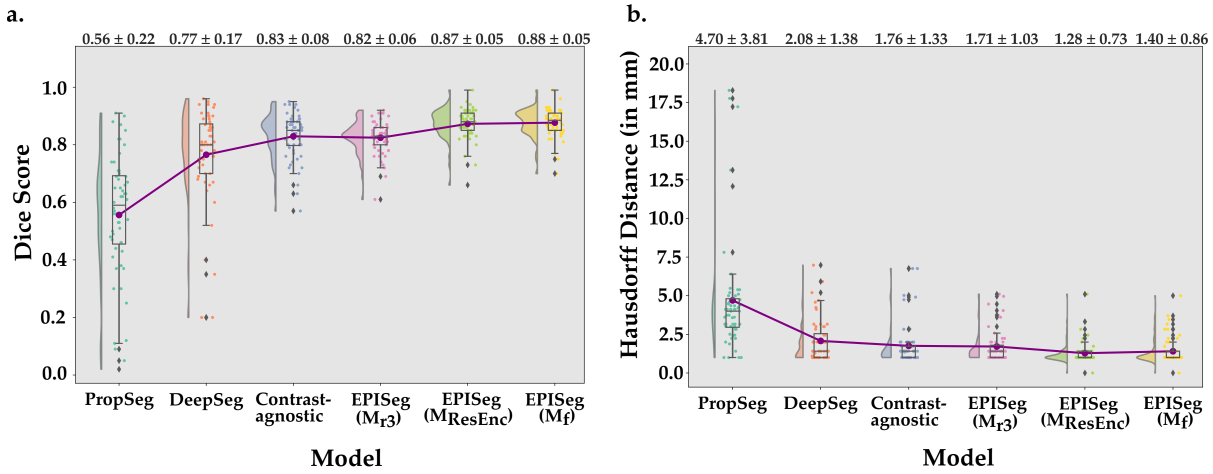

Figure 4.4 **Raincloud plots from method comparison.** 4a. The DS were computed between the ground truth and predictions. The raincloud plots show the DS distributions of the different models. The red dots and the line connect the mean values of the DS of the respective models. The mean DS along with the standard deviation is printed on the top of the plot. 4b. In a similar way to 4a., it shows the HD95 for the different models.

4.6.3 Comparison with other methods

We compared EPISeg with other methods, namely, PropSeg, DeepSeg and Contrast-agnostic (from SCT v6.3: `https://github.com/spinalcordtoolbox/spinalcordtoolbox/releases/tag/6.3`). PropSeg does not use DL for SC segmentation and has been tested only on T1, T2 and T2* contrasts, DeepSeg uses DL but no EPI data were used during the training. Similarly, Contrast-agnostic did not use gradient-echo EPI data during training. However, both DeepSeg and Contrast-agnostic can segment EPI data. Our motive to compare our results with these methods was the fact that they are considered state-of-the-art, fully automatic 3D SC segmentation methods.

In Table 4.3, the various model performances on D_{test} are presented. The proposed model, EPISeg, exhibits superior performance in comparison to the alternative baselines. Both PropSeg and DeepSeg depict subpar performance when dealing with EPI data. Numerous instances exist where these models struggle to accurately segment the SC. PropSeg, for exam-

Table 4.3 Comparison of SC segmentation methods on D_{test}

Model	**DS ± S.D.**	**HD95 ± S.D. (in mm)**
Propseg	0.56 ± 0.22	4.70 ± 3.81
Deepseg	0.77 ± 0.17	2.08 ± 1.38
Contrast-agnostic	0.83 ± 0.08	1.76 ± 1.33
EPISeg (M_{r3})	0.82 ± 0.06	1.73 ± 1.03
EPISeg (M_f)	**0.88 ± 0.05**	1.41 ± 0.86
EPISeg (M_{ResEnc})	0.87 ± 0.05	**1.28 ± 0.73**

ple, faces difficulties in identifying the SC edges in many scenarios, resulting in segmentation failure. Conversely, Deepseg initially locates the center of mass before proceeding with SC segmentation, a task it struggles with in EPI SC segmentation. The Contrast-agnostic model demonstrates satisfactory but lower performance on our dataset due to its exposure to diverse SC contrasts during the training process.

Figure 4.4 presents raincloud plots comparing the performance of different methods on D_{test}. The half-violin plots illustrate the distribution of DS, while individual points represent the performance of each data point in D_{test}. EPISeg's DS are tightly clustered around a single point, which coincides with both the mean and median of the scores. This concentration indicates that EPISeg delivers consistent performance without outliers. When it comes to comparison of HD95 in Figure 4.5, we observe a similar pattern as in Figure 4.4. To highlight the strong performance of EPISeg, we see a point where the HD95 value is 0.0mm which, means that 95% or more surface points had 0.0mm error, indicating a perfect segmentation performance.

4.6.4 Qualitative assessment

We evaluated our model's performance on D_{test} through a qualitative assessment involving two raters with 6 and 18 years of experience of SC MRI experience, respectively. These raters meticulously examined the segmentation results for each slice, as documented in the QC report generated by the sct_qc functionality in SCT. Figure 4.5 showcases the performance of various models on selected axial slices, highlighting the quantitative performance of each. The participants represented in the figure include healthy controls as well as patients with degenerative cervical myelopathy and fibromyalgia.

In our evaluations, we observed that PropSeg frequently under-segments and sometimes fails to segment the SC entirely, especially when the SC is not clearly distinguishable from the cerebrospinal fluid. DeepSeg performs slightly better but still displays inconsistencies in

SC segmentation. The Contrast-agnostic model offers more consistent segmentation boundaries but often over-segments the SC. EPISeg demonstrates superior performance, closely matching the ground truth. For instance, EPISeg successfully segmented the SC in the sub-ZurLumbC01, a dataset with low-quality lumbar cord images that were not included in the training set.

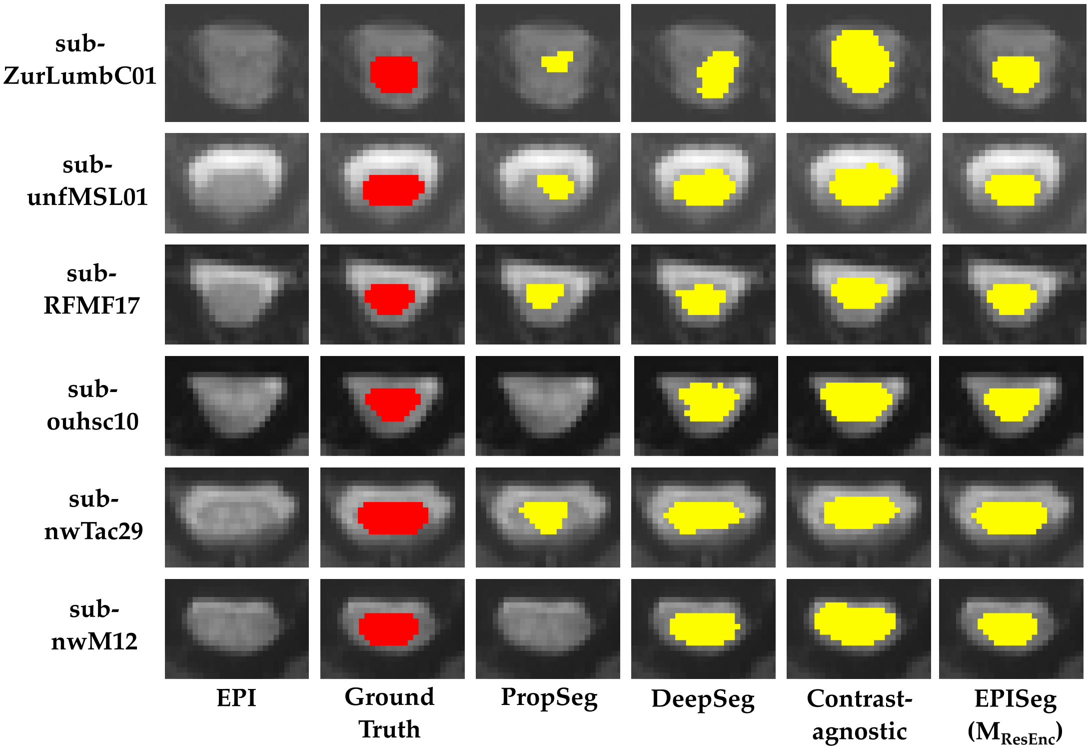

Figure 4.5 **Visual assessment.** Comparison of EPISeg (M_{ResEnc}) with other automatic SC segmentation methods. The colored overlay shows the SC on the images. We can observe that EPISeg provides the most precise segmentations.

4.7 Conclusion

We provide an automatic SC segmentation method for gradient-echo EPI. We demonstrate the performance of our method on various spinal and also corticospinal fMRI datasets (including samples from clinical populations) from different sites with different acquisition parameters. The dataset is publicly available on OpenNeuro. This reliable segmentation method will provide a strong basis for fMRI data preprocessing, particularly for spatial normalization which is necessary for obtaining reliable group-level results. Our method is readily available as a part of SCT v6.3 or higher.

Data and Code Availability

All the data used for this study is open-sourced and can be found at: `https://openneuro.org/datasets/ds005143/versions/1.3.0`

All the scripts, instructions to reproduce results and the lastest model weights can be found here: https://github.com/sct-pipeline/fmri-segmentation/releases/tag/v0.2. This method is integrated into SCT v6.3 and can be used by using the *sct_deepseg* command with the *-task seg_sc_epi* parameter.

4.8 Author Contributions

Rohan Banerjee: Data Curation, Methodology, Formal Analysis, Software, Validation, Writing – Original Draft, Writing – Review Editing, Visualization; Merve Kaptan: Data Curation, Validation, Methodology, Writing – Original Draft, Writing – Review Editing; Kenneth A. Weber II: Supervision, Writing – Review Editing, Conceptualization, Funding Acquisition; Benjamin De Leener: Supervision, Methodology, Writing – Review Editing, Conceptualization, Funding Acquisition; Julien Cohen-Adad: Supervision, Methodology, Writing – Review Editing, Conceptualization, Funding Acquisition; Alexandra Tinnermann: Data Contribution; Ali Khatibi: Data Contribution; Alice Dabbagh: Data Contribution; Christian Büchel: Data Contribution; Christian W. Kündig: Data Contribution; Christine S.W. Law: Data Contribution; Dario Pfyffer: Data Contribution, Writing – Review Editing; David J. Lythgoe: Data Contribution, Writing – Review Editing; Dimitra Tsivaka: Data Contribution; Dimitri Van De Ville: Data Contribution; Falk Eippert: Data Contribution, Writing – Review Editing; Fauziyya Muhammad: Data Contribution; Gary H. Glover: Funding Acquisition, Data Contribution; Gergely David: Data Contribution, Writing – Review Editing; Grace Haynes: Data Contribution; Jan Haaker: Data Contribution, Writing – Review Editing; Jonathan C. W. Brooks: Data Contribution, Writing – Review Editing; Jürgen Finsterbusch: Data Contribution, Writing – Review Editing; Katherine T. Martucci: Data Contribution, Funding Acquisition, Writing – Review Editing; Kimberly J. Hemmerling: Data Contribution, Mahdi Mobarak-Abadi: Data Contribution; Mark A. Hoggarth:Data Contribution; Matthew A. Howard: Data Contribution; Molly G. Bright: Data Contribution; Nawal Kinany: Data Contribution; Olivia S. Kowalczyk: Data Contribution; Patrick Freund: Data Contribution; Robert L. Barry: Data Contribution, Funding Acquisition, Writing – Review Editing; Sean Mackey: Funding Acquisition, Data Contribution; Simon Schading: Data Contribution; Shahabeddin Vahdat: Data Contribution; Todd Parish: Data Contribution; Yufen Chen: Data Contribution; Véronique Marchand-Pauvert: Data Contribution;

Zachary A. Smith: Data Contribution

4.9 Funding

This study was funded by the Canada Research Chair in Quantitative Magnetic Resonance Imaging [CRC-2020-00179], the Canadian Institute of Health Research [PJT-190258], the Canada Foundation for Innovation [32454, 34824], the Fonds de Recherche du Québec - Santé [322736, 324636], the Natural Sciences and Engineering Research Council of Canada [RGPIN-2019-07244], the Canada First Research Excellence Fund (IVADO and TransMedTech), the Quebec BioImaging Network [5886, 35450], INSPIRED (Spinal Research, UK; Wings for Life, Austria; Craig H. Neilsen Foundation, USA), Mila - Tech Transfer Funding Program, the National Institute of Neurological Disorders and Stroke (K23NS104211, R01NS109450, K24NS126781, R01NS133305, and R01NS128478), the National Institute on Drug Abuse (K99/R00DA040154), the National Center for Complementary and Integrative Health (F32AT007800), and the National Institute of Biomedical Imaging and Bioengineering (R01EB027779 and R21EB031211), the Swiss National Science Foundation (SNSF; 32003B_204934). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

4.10 Declaration of Competing Interests

Since January 2024, Dr. Barry has been employed by the National Institute of Biomedical Imaging and Bioengineering at the National Institutes of Health. This work was co-authored by Robert Barry in his personal capacity. The opinions expressed in this study are his own and do not necessarily reflect the views of the National Institutes of Health, the Department of Health and Human Services, or the United States government.

The other authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CHAPTER 5 DISCUSSION

In this Master's thesis, we introduce EPISeg, a DL-based method for automatic SC segmentation from gradient-echo EPI images, which are widely used in (fMRI). EPISeg is robust to heterogeneous resolutions, scanner settings, and motion artifacts. Consequently, it addresses challenges that have historically hindered SC segmentation on EPI data, requiring labor-intensive and bias-prone manual interventions.

In order to develop a robust EPI segmentation model, the first necessary step was to create a dataset that would provide a DL model with enough context to be able to segment the SC in a low-quality image setting. We curated a large multi-site EPI SC dataset consisting of spinal and corticospinal acquisitions, arranged it as per the BIDS convention, and made it publicly available in OpenNeuro - `https://openneuro.org/datasets/ds005143/versions/1.3.0`.

While creating the dataset, we observed that the provided ground truth was erroneous in some cases (e.g., over-segmented or under-segmented). This led us to the decision to train EPISeg sequentially. We used good-quality segmentations to train a strong baseline model which provided high-quality segmentations before successive models. At each model training step, a pool of best predictions was chosen along with manual correction wherever required, including verification by an expert. We obtained our final model using these steps iteratively, which was able to segment the SC reliably. We demonstrated the robustness of our model against low spatial resolution of EPIs, susceptibility artifacts, signal drop-outs, ghosting, and motion-related distortions and eliminated the need for manual interventions.

In terms of SC segmentation directly on EPI data, only [45] provides a semi-automatic method. This method consists of multiple preprocessing steps before the SC was segmented using k-means clustering. The authors claim that their method's performance efficiency declines with the increase in the distortion present in the image. Even though we cannot reproduce this method for comparison due to the lack of provided details regarding the parameters used, we show that, in addition to being completely automatic, EPIseg is highly robust to distortions present in EPI data.

Our method, EPISeg, demonstrates superior accuracy and robustness in spinal cord segmentation compared to state-of-the-art models like PropSeg, DeepSeg, and Contrast-agnostic segmentation. It generalizes well across different SC conditions and handles out-of-distribution data effectively. EPISeg achieves a mean DS of 0.88 with a standard deviation of 0.05, reflecting its consistent performance. PropSeg performs poorly as it was primarily optimized for T1 and T2 contrasts and is not well-suited for EPI data. Similarly, DeepSeg and the contrast-agnostic model, which were not specifically trained for spinal cord segmentation on EPI images, lack sufficient context to accurately handle the unique characteristics of EPI data. Although the Contrast-agnostic model performs better than PropSeg and DeepSeg due to its broader training dataset, it still falls short of EPISeg's performance.

To potentially circumvent the need for segmenting the EPI for registration purposes, a high-resolution anatomical image can be acquired with the same field-of-view as the EPI images (in a way that registration between this image and the EPI is not necessary). However, this approach has also drawbacks. Mainly, the acquisition of a high-resolution image would require additional scan time. Especially considering the movement of a participant within and across multiple functional runs, it may be even necessary to acquire one high-resolution image which would significantly increase scan time.

Despite the superior performance of EPISeg for segmentation of spinal cord fMRI data against other automated methods, several limitations of the current work must be considered. Although this study includes a wide range of data from multiple sites and participants including data from different patient populations, it does not cover all possible data domains. Variations in acquisition parameters and spinal cord conditions across different populations may not be fully represented here, possibly limiting the generalizability of our findings. Additionally, our method has solely been tested on data acquired at 3T. Segmentation performance and accuracy could vary across other field strengths, due to the fact that, at lower (1.5T) and higher (7T) fields, differences in signal-to-noise ratio, contrast, and spatial resolution may influence the quality of the EPI data.

EPISeg here is optimized for EPI sequences as they are the most commonly used technique for fMRI acquisitions. However, spinal cord fMRI can also be performed using other sequence types (such as T2*-weighted fast gradient echo acquisitions). The performance of EPIseg on other acquisition types has not been evaluated in this study. Finally, although we gathered data from various sites that have employed different motion correction techniques, it is important to remember that effective motion correction is critical to the success of spinal cord fMRI segmentation, given the susceptibility of this region to motion artifacts from respiration, pulsation, and participant movement. Inadequate motion correction could impact the accuracy of segmentation by obscuring spinal cord boundaries. Future work should aim to broaden the data domains represented, including a wider range of acquisition parameters, populations, and magnetic field strengths (1.5T, 3T, and 7T). Moreover, validation of EPIseg on GRE T2*-weighted sequences and improvements in motion correction strategies will be necessary to enhance segmentation robustness and ensure that it performs effectively across diverse imaging conditions.

CHAPTER 6 CONCLUSION

This thesis, culminating in the development of EPISeg, was deeply influenced by the principles of open science and reproducibility fostered within the multi-disciplinary environment at NeuroPoly, under the guidance of Prof. Benjamin De Leener and Prof. Julien Cohen-Adad. The project directly addressed a need identified through feedback from the SC fMRI community: the necessity for a standardized, automated SC segmentation procedure applicable to challenging EPI data, alongside the creation of a large-scale, publicly available dataset to support reproducible research in this domain. EPISeg, presented herein, successfully fulfills these requirements by providing a robust DL-based solution for SC segmentation on EPI images, accompanied by a comprehensive multi-center dataset shared via OpenNeuro.

The practical utility of EPISeg is already being realized, with collaborators beginning to integrate the tool into their analysis pipelines via its implementation in SCT, providing positive initial feedback on its performance. Looking ahead, several avenues exist to further enhance and build upon this work.

1. Exploring initialization with pre-trained model weights holds significant promise. Leveraging large corpora of unlabeled or labeled SC data for pre-training, despite contrasting views on the optimal strategy (e.g., supervised pretraining [109] vs. self-supervised pre-training [110]), has generally proven beneficial for MIS and warrants investigation for potentially improving EPISeg's performance or data efficiency.

2. It would be valuable to extend the work of Hoggarth et al. [17] by systematically evaluating the impact of using an automated method like EPISeg, compared to manual contouring, on the downstream processes of fMRI co-registration and interpretation. This would provide crucial insights into the practical benefits of automation for analysis reproducibility.

3. Enhancing EPISeg to produce uncertainty maps alongside segmentations could significantly increase its clinical utility, providing clinicians with a measure of confidence in the automated results to inform their decision-making when utilizing these segmentations.

REFERENCES

[1] C. Sprenger *et al.*, “Attention modulates spinal cord responses to pain,” *Current Biology*, vol. 22, pp. 1019–1022, 2012. [Online]. Available: https://api.semanticscholar.org/CorpusID:17142233

[2] C. Sprenger, J. Finsterbusch, and C. Büchel, “Spinal cord–midbrain functional connectivity is related to perceived pain intensity: A combined spino-cortical fmri study,” *The Journal of Neuroscience*, vol. 35, pp. 4248 – 4257, 2015. [Online]. Available: https://api.semanticscholar.org/CorpusID:26721434

[3] K. A. Weber *et al.*, “Lateralization of cervical spinal cord activity during an isometric upper extremity motor task with functional magnetic resonance imaging,” *NeuroImage*, vol. 125, pp. 233–243, 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:29015234

[4] N. Kinany *et al.*, “Functional imaging of rostrocaudal spinal activity during upper limb motor tasks,” *NeuroImage*, vol. 200, pp. 590–600, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:156056105

[5] R. L. Barry *et al.*, “Resting state functional connectivity in the human spinal cord,” *eLife*, vol. 3, 2014. [Online]. Available: https://api.semanticscholar.org/CorpusID:6070136

[6] Y. Kong *et al.*, “Intrinsically organized resting state networks in the human spinal cord,” *Proceedings of the National Academy of Sciences*, vol. 111, pp. 18 067 – 18 072, 2014. [Online]. Available: https://api.semanticscholar.org/CorpusID:7718279

[7] N. Kinany *et al.*, “Dynamic functional connectivity of resting-state spinal cord fmri reveals fine-grained intrinsic architecture,” *Neuron*, vol. 108, pp. 424–435.e4, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:221567643

[8] ——, “Spinal cord fmri: A new window into the central nervous system,” *The Neuroscientist*, vol. 29, pp. 715 – 731, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:250489772

[9] A. R. Martin *et al.*, “Translating state-of-the-art spinal cord mri techniques to clinical use: A systematic review of clinical studies utilizing dti, mt, mwf, mrs,

and fmri," *NeuroImage : Clinical*, vol. 10, pp. 192 – 238, 2015. [Online]. Available: https://api.semanticscholar.org/CorpusID:465694

[10] J. M. Powers, G. Ioachim, and P. W. Stroman, "Ten key insights into the use of spinal cord fmri," *Brain Sciences*, vol. 8, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:52185170

[11] P. W. Stroman *et al.*, "Functional magnetic resonance imaging of the human cervical spinal cord with stimulation of different sensory dermatomes." *Magnetic resonance imaging*, vol. 20 1, pp. 1–6, 2002. [Online]. Available: https://api.semanticscholar.org/CorpusID:28664443

[12] ——, "Pain processing in the human brainstem and spinal cord before, during, and after the application of noxious heat stimuli," *PAIN*, vol. 159, p. 2012–2020, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:49210063

[13] D. W. Cadotte *et al.*, "Plasticity of the injured human spinal cord: Insights revealed by spinal cord functional mri," *PLoS ONE*, vol. 7, 2012. [Online]. Available: https://api.semanticscholar.org/CorpusID:17774297

[14] J. Cohen-Adad, "Functional magnetic resonance imaging of the spinal cord: Current status and future developments." *Seminars in ultrasound, CT, and MR*, vol. 38 2, pp. 176–186, 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:40687893

[15] M. Fratini *et al.*, "On the impact of physiological noise in spinal cord functional mri," *Journal of Magnetic Resonance Imaging*, vol. 40, 2014. [Online]. Available: https://api.semanticscholar.org/CorpusID:2063624

[16] R. Banerjee *et al.*, "Episeg: Automated segmentation of the spinal cord on echo planar images using open-access multi-center data," *bioRxiv*, 2025. [Online]. Available: https://api.semanticscholar.org/CorpusID:275458889

[17] M. A. Hoggarth *et al.*, "Effects of variability in manually contoured spinal cord masks on fmri co-registration and interpretation," *Frontiers in Neurology*, vol. 13, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:247780741

[18] J. Cohen-Adad *et al.*, "Generic acquisition protocol for quantitative mri of the spinal cord," *Nature Protocols*, vol. 16, pp. 4611 – 4632, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:237147935

[19] C. Landelle *et al.*, "Investigating the human spinal sensorimotor pathways through functional magnetic resonance imaging," *NeuroImage*, vol. 245, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:240277132

[20] C. Watson and G. Kayalioglu, "The organization of the spinal cord," 2009. [Online]. Available: https://api.semanticscholar.org/CorpusID:68446762

[21] C. S. Ahuja *et al.*, "Traumatic spinal cord injury," *Nature Reviews Disease Primers*, vol. 3, 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:24244894

[22] R. Bitar *et al.*, "Mr pulse sequences: what every radiologist wants to know but is afraid to ask." *Radiographics : a review publication of the Radiological Society of North America, Inc*, vol. 26 2, pp. 513–37, 2006. [Online]. Available: https://api.semanticscholar.org/CorpusID:31134167

[23] R. B. Buxton, "The physics of functional magnetic resonance imaging (fmri)," *Reports on Progress in Physics*, vol. 76, 2013. [Online]. Available: https://api.semanticscholar.org/CorpusID:33748831

[24] S. Ogawa *et al.*, "Brain magnetic resonance imaging with contrast dependent on blood oxygenation." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 87 24, pp. 9868–72, 1990. [Online]. Available: https://api.semanticscholar.org/CorpusID:8183987

[25] R. A. Pooley, "Aapm/rsna physics tutorial for residents: fundamental physics of mr imaging." *Radiographics : a review publication of the Radiological Society of North America, Inc*, vol. 25 4, pp. 1087–99, 2005. [Online]. Available: https://api.semanticscholar.org/CorpusID:21823453

[26] H. Matsuzaki *et al.*, "The origin and significance of spinal cord pulsation," *Spinal Cord*, vol. 34, pp. 422–426, 1996. [Online]. Available: https://api.semanticscholar.org/CorpusID:12947245

[27] T. Verma and J. Cohen-Adad, "Effect of respiration on the b0 field in the human spinal cord at 3t," *Magnetic Resonance in Medicine*, vol. 72, 2014. [Online]. Available: https://api.semanticscholar.org/CorpusID:46338985

[28] S. Lévy *et al.*, "White matter atlas of the human spinal cord with estimation of partial volume effect," *NeuroImage*, vol. 119, pp. 262–271, 2015. [Online]. Available: https://api.semanticscholar.org/CorpusID:8163596

[29] P. W. Stroman and L. Ryner, “Functional mri of motor and sensory activation in the human spinal cord.” *Magnetic resonance imaging*, vol. 19 1, pp. 27–32, 2001. [Online]. Available: https://api.semanticscholar.org/CorpusID:36108778

[30] C. A. M. Wheeler-Kingshott *et al.*, “The current state-of-the-art of spinal cord imaging: Applications,” *NeuroImage*, vol. 84, pp. 1082–1093, 2014. [Online]. Available: https://api.semanticscholar.org/CorpusID:12191293

[31] S. Vahdat *et al.*, “Simultaneous brain–cervical cord fmri reveals intrinsic spinal cord plasticity during motor sequence learning,” *PLoS Biology*, vol. 13, 2015. [Online]. Available: https://api.semanticscholar.org/CorpusID:14823503

[32] B. De Leener *et al.*, “Sct: Spinal cord toolbox, an open-source software for processing spinal cord mri data,” *NeuroImage*, vol. 145, pp. 24–43, 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:10981848

[33] K. J. Gorgolewski *et al.*, “The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments,” *Scientific Data*, vol. 3, 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:795781

[34] ——, “Bids apps: Improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods,” *PLoS Computational Biology*, vol. 13, 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:2442453

[35] S. Bédard *et al.*, “Towards contrast-agnostic soft segmentation of the spinal cord,” *ArXiv*, vol. abs/2310.15402, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:264438952

[36] B. De Leener *et al.*, “Propseg: automatic spinal cord segmentation method for mr images using propagated deformable models,” *F1000Research*, vol. 5, 2014. [Online]. Available: https://api.semanticscholar.org/CorpusID:74119587

[37] C. Blanc *et al.*, “Combining propseg and a convolutional neural network for automatic spinal cord segmentation in pediatric populations and patients with spinal cord injury,” *International Journal of Imaging Systems and Technology*, vol. 33, pp. 1396 – 1405, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:256607698

[38] C. Gros *et al.*, “Automatic segmentation of the spinal cord and intramedullary multiple sclerosis lesions with convolutional neural networks,” *NeuroImage*, vol. 184, pp. 901–915, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:21699394

[39] E. N. Karthik *et al.*, “Sciseg: Automatic segmentation of t2-weighted intramedullary lesions in spinal cord injury,” *medRxiv*, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:266743496

[40] M. A. Horsfield *et al.*, “Rapid semi-automatic segmentation of the spinal cord from magnetic resonance images: Application in multiple sclerosis,” *NeuroImage*, vol. 50, pp. 446–455, 2010. [Online]. Available: https://api.semanticscholar.org/CorpusID:34258683

[41] M. Hohenhaus *et al.*, “Quantification of cervical spinal stenosis by automated 3D MRI segmentation of spinal cord and cerebrospinal fluid space,” *Spinal Cord*, Apr. 2024.

[42] C. Tsagkas *et al.*, “Fully automatic method for reliable spinal cord compartment segmentation in multiple sclerosis,” *AJNR Am. J. Neuroradiol.*, vol. 44, no. 2, pp. 218–227, Feb. 2023.

[43] H. Toufani, A. Vard, and I. Adibi, “A pipeline to quantify spinal cord atrophy with deep learning: Application to differentiation of MS and NMOSD patients,” *Phys. Med.*, vol. 89, pp. 51–62, Sep. 2021.

[44] C. Lukas *et al.*, “Quantification of cervical cord cross-sectional area: Which acquisition, vertebra level, and analysis software? a multicenter repeatability study on a traveling healthy volunteer,” *Front. Neurol.*, vol. 12, p. 693333, Aug. 2021.

[45] G. Giulietti *et al.*, “Semiautomated segmentation of the human spine based on echoplanar images.” *Magnetic resonance imaging*, vol. 29 10, pp. 1429–36, 2011. [Online]. Available: https://api.semanticscholar.org/CorpusID:2224041

[46] J. Valosek and J. Cohen-Adad, “Reproducible spinal cord quantitative mri analysis with the spinal cord toolbox,” *Magnetic Resonance in Medical Sciences*, vol. 23, pp. 307 – 315, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:268382554

[47] L. Luchetti *et al.*, “Evaluation of cervical spinal cord atrophy using a modified siena approach,” *NeuroImage*, vol. 298, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:271672281

[48] B. D. Leener *et al.*, “Segmentation of the human spinal cord,” *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 29, pp. 125–153, 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:10316667

[49] N. A. Losseff *et al.*, “Spinal cord atrophy and disability in multiple sclerosis. a new reproducible and sensitive mri method with potential to monitor disease progression.” *Brain : a journal of neurology*, vol. 119 (Pt 3), pp. 701–8, 1996. [Online]. Available: https://api.semanticscholar.org/CorpusID:7305421

[50] O. Coulon *et al.*, “Quantification of spinal cord atrophy from magnetic resonance images via a b-spline active surface model,” *Magnetic Resonance in Medicine*, vol. 47, 2002. [Online]. Available: https://api.semanticscholar.org/CorpusID:138847

[51] B. De Leener *et al.*, “Pam50: Unbiased multimodal template of the brainstem and spinal cord aligned with the icbm152 space,” *NeuroImage*, vol. 165, pp. 170–179, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:7337168

[52] Y. LeCun *et al.*, “Backpropagation applied to handwritten zip code recognition,” *Neural Computation*, vol. 1, pp. 541–551, 1989. [Online]. Available: https://api.semanticscholar.org/CorpusID:41312633

[53] J. Gu *et al.*, “Recent advances in convolutional neural networks,” *Pattern Recognit.*, vol. 77, pp. 354–377, 2015. [Online]. Available: https://api.semanticscholar.org/CorpusID:3879949

[54] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, pp. 84 – 90, 2012. [Online]. Available: https://api.semanticscholar.org/CorpusID:195908774

[55] G. J. S. Litjens *et al.*, “A survey on deep learning in medical image analysis,” *Medical image analysis*, vol. 42, pp. 60–88, 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:2088679

[56] D. Shen, G. Wu, and H.-I. Suk, “Deep learning in medical image analysis.” *Annual review of biomedical engineering*, vol. 19, pp. 221–248, 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:207614217

[57] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in *International Conference on Machine Learning*, 2010. [Online]. Available: https://api.semanticscholar.org/CorpusID:15539264

[58] M. D. Zeiler and R. Fergus, “Visualizing and understanding convolutional networks,” *ArXiv*, vol. abs/1311.2901, 2013. [Online]. Available: https://api.semanticscholar.org/CorpusID:3960646

[59] E. Shelhamer, J. Long, and T. Darrell, “Fully convolutional networks for semantic segmentation,” *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431–3440, 2014. [Online]. Available: https://api.semanticscholar.org/CorpusID:1629541

[60] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” *ArXiv*, vol. abs/1505.04597, 2015. [Online]. Available: https://api.semanticscholar.org/CorpusID:3719281

[61] N. A. Siddique *et al.*, “U-net and its variants for medical image segmentation: A review of theory and applications,” *IEEE Access*, vol. 9, pp. 82 031–82 057, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:226236779

[62] B. Kayalibay, G. Jensen, and P. van der Smagt, “Cnn-based segmentation of medical imaging data,” *ArXiv*, vol. abs/1701.03056, 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:17581107

[63] Ö. Çiçek *et al.*, “3d u-net: Learning dense volumetric segmentation from sparse annotation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:2164893

[64] L. R. Dice, “Measures of the amount of ecologic association between species,” *Ecology*, vol. 26, pp. 297–302, 1945. [Online]. Available: https://api.semanticscholar.org/CorpusID:53335638

[65] D. P. Huttenlocher, G. A. Klanderman, and W. Rucklidge, “Comparing images using the hausdorff distance,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, pp. 850–863, 1993. [Online]. Available: https://api.semanticscholar.org/CorpusID:8027136

[66] L. Yang *et al.*, “Suggestive annotation: A deep active learning framework for biomedical image segmentation,” *ArXiv*, vol. abs/1706.04737, 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:22622013

[67] S. Budd, E. C. Robinson, and B. Kainz, “A survey on active learning and human-in-the-loop deep learning for medical image analysis,” *Medical image analysis*, vol. 71, p. 102062, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:203837078

[68] A. Holzinger, “Interactive machine learning for health informatics: when do we need the human-in-the-loop?” *Brain Informatics*, vol. 3, pp. 119 – 131, 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:4649427

[69] B. Settles, “Active learning literature survey,” 2009. [Online]. Available: https://api.semanticscholar.org/CorpusID:324600

[70] D. D. Lewis and W. A. Gale, “A sequential algorithm for training text classifiers,” in *Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1994. [Online]. Available: https://api.semanticscholar.org/CorpusID:915058

[71] H. S. Seung, M. Opper, and H. Sompolinsky, “Query by committee,” in *Annual Conference Computational Learning Theory*, 1992. [Online]. Available: https://api.semanticscholar.org/CorpusID:7869993

[72] M. McCloskey and N. J. Cohen, “Catastrophic interference in connectionist networks: The sequential learning problem,” *Psychology of Learning and Motivation*, vol. 24, pp. 109–165, 1989. [Online]. Available: https://api.semanticscholar.org/CorpusID:61019113

[73] R. Ratcliff, “Connectionist models of recognition memory: constraints imposed by learning and forgetting functions.” *Psychological review*, vol. 97 2, pp. 285–308, 1990. [Online]. Available: https://api.semanticscholar.org/CorpusID:18556305

[74] G. Haynes *et al.*, “The current state of spinal cord functional magnetic resonance imaging and its application in clinical research,” *Journal of Neuroimaging*, vol. 33, pp. 877 – 888, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:262147906

[75] L. M. Chen *et al.*, “Injury alters intrinsic functional connectivity within the primate spinal cord,” *Proceedings of the National Academy of Sciences*, vol. 112, pp. 5991 – 5996, 2015. [Online]. Available: https://api.semanticscholar.org/CorpusID:20499101

[76] C. Landelle *et al.*, “Altered spinal cord functional connectivity associated with parkinson’s disease progression,” *Movement Disorders*, vol. 38, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:257077183

[77] R. Staud *et al.*, “Spinal cord neural activity of fibromyalgia patients and healthy controls during temporal summation of pain: An fmri study.” *Journal of*

neurophysiology, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:237214626

[78] B. N. Conrad *et al.*, "Multiple sclerosis lesions affect intrinsic functional connectivity of the spinal cord," *Brain*, vol. 141, p. 1650–1664, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:4931193

[79] P. W. Stroman *et al.*, "The current state-of-the-art of spinal cord imaging: Methods," *NeuroImage*, vol. 84, pp. 1070–1081, 2014. [Online]. Available: https://api.semanticscholar.org/CorpusID:14443462

[80] T.-L. Wu *et al.*, "Intrinsic functional architecture of the non-human primate spinal cord derived from fmri and electrophysiology," *Nature Communications*, vol. 10, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:85567049

[81] A. Tinnermann, C. Sprenger, and C. Büchel, "Opioid analgesia alters corticospinal coupling along the descending pain system in healthy participants," *eLife*, vol. 11, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:248390204

[82] K. A. Weber *et al.*, "Assessing the spatial distribution of cervical spinal cord activity during tactile stimulation of the upper extremity in humans with functional magnetic resonance imaging," *NeuroImage*, vol. 217, pp. 116 905 – 116 905, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:218516504

[83] A. Sengupta *et al.*, "Characteristic bold signals are detectable in white matter of the spinal cord at rest and after a stimulus." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 121 22, p. e2316117121, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:269974582

[84] K. J. Hemmerling *et al.*, "Spatial distribution of hand-grasp motor task activity in spinal cord functional magnetic resonance imaging," *Human Brain Mapping*, vol. 44, pp. 5567 – 5581, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:261073944

[85] A. Khatibi *et al.*, "Brain-spinal cord interaction in long-term motor sequence learning in human: An fmri study," *NeuroImage*, vol. 253, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:247586476

[86] K. Weber *et al.*, "Functional magnetic resonance imaging of the cervical spinal cord during thermal stimulation across consecutive runs," *NeuroImage*, vol. 143, 09 2016.

[87] K. A. Weber *et al.*, "Thermal stimulation alters cervical spinal cord functional connectivity in humans," *Neuroscience*, vol. 369, pp. 40–50, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:23340290

[88] C. W. Kündig *et al.*, "Functional magnetic resonance imaging of the lumbosacral cord during a lower extremity motor task," *Imaging Neuroscience*, vol. 2, pp. 1–19, 07 2024. [Online]. Available: https://doi.org/10.1162/imag_a_00227

[89] M. K. Stehling, R. Turner, and P. Mansfield, "Echo-planar imaging: magnetic resonance imaging in a fraction of a second." *Science*, vol. 254 5028, pp. 43–50, 1991. [Online]. Available: https://api.semanticscholar.org/CorpusID:26600241

[90] F. Giove *et al.*, "Issues about the fmri of the human spinal cord." *Magnetic resonance imaging*, vol. 22 10, pp. 1505–16, 2004. [Online]. Available: https://api.semanticscholar.org/CorpusID:9559516

[91] V. S. Fonov *et al.*, "Framework for integrated mri average of the spinal cord white and gray matter: The mni–poly–amu template," *NeuroImage*, vol. 102, pp. 817–827, 2014. [Online]. Available: https://api.semanticscholar.org/CorpusID:1577725

[92] M. Azzarito *et al.*, "Simultaneous voxel-wise analysis of brain and spinal cord morphometry and microstructure within the spm framework," *Human Brain Mapping*, vol. 42, pp. 220 – 232, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:222157395

[93] M. Taso *et al.*, "Construction of an in vivo human spinal cord atlas based on high-resolution mr images at cervical and thoracic levels: preliminary results," *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 27, pp. 257–267, 2014. [Online]. Available: https://api.semanticscholar.org/CorpusID:24635366

[94] R. L. Bosma and P. W. Stroman, "Assessment of data acquisition parameters, and analysis techniques for noise reduction in spinal cord fMRI data," vol. 32, no. 5, pp. 473–481. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0730725X14000113

[95] P. W. Stroman, C. R. Figley, and C. M. Cahill, "Spatial normalization, bulk motion correction and coregistration for functional magnetic resonance imaging of the human cervical spinal cord and brainstem." *Magnetic resonance imaging*, vol. 26 6, pp. 809–14, 2008. [Online]. Available: https://api.semanticscholar.org/CorpusID:207027755

[96] M. Kaptan *et al.*, “Automated slice-specific z-shimming for functional magnetic resonance imaging of the human spinal cord,” *Human Brain Mapping*, vol. 43, pp. 5389 – 5407, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:251399718

[97] ——, “Reliability of resting-state functional connectivity in the human spinal cord: Assessing the impact of distinct noise sources,” *Neuroimage*, vol. 275, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:255126115

[98] O. S. Kowalczyk *et al.*, “Spinal fmri demonstrates segmental organisation of functionally connected networks in the cervical spinal cord: A test–retest reliability study,” *Human Brain Mapping*, vol. 45, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:267380960

[99] M. Kaptan *et al.*, “Exploring corticospinal functional connectome using resting-state functional magnetic resonance imaging,” *The Journal of Pain*, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:257974875

[100] K. T. Martucci, K. A. Weber, and S. C. Mackey, “Spinal cord resting state activity in individuals with fibromyalgia who take opioids,” *Frontiers in Neurology*, vol. 12, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:236900175

[101] ——, “Altered cervical spinal cord resting-state activity in fibromyalgia,” *Arthritis & Rheumatology*, vol. 71, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:52911043

[102] M. Mobarak-Abadi *et al.*, “DeepRetroMoCo: deep neural network-based retrospective motion correction algorithm for spinal cord functional MRI,” vol. 15, publisher: Frontiers. [Online]. Available: https://www.frontiersin.org/journals/psychiatry/articles/10.3389/fpsyt.2024.1323109/full

[103] A. Tinnermann, C. Büchel, and J. Haaker, “Observation of others’ painful heat stimulation involves responses in the spinal cord,” *Science Advances*, vol. 7, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:232480874

[104] A. Dabbagh *et al.*, “Reliability of task-based fMRI in the dorsal horn of the human spinal cord,” vol. 2, pp. 1–27. [Online]. Available: https://doi.org/10.1162/imag_a_00273

[105] J. Finsterbusch, F. Eippert, and C. Büchel, “Single, slice-specific z-shim gradient pulses improve t2*-weighted imaging of the spinal cord,” *NeuroImage*, vol. 59, pp. 2307–2315, 2012. [Online]. Available: https://api.semanticscholar.org/CorpusID:5259447

[106] F. Isensee *et al.*, “nnu-net: a self-configuring method for deep learning-based biomedical image segmentation,” *Nature Methods*, vol. 18, pp. 203 – 211, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:227947847

[107] ——, “nnu-net revisited: A call for rigorous validation in 3d medical image segmentation,” *ArXiv*, vol. abs/2404.09556, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:269148560

[108] L. Maier-Hein *et al.*, “Metrics reloaded: recommendations for image analysis validation.” *Nature methods*, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:249375598

[109] W. Li, A. L. Yuille, and Z. Zhou, “How well do supervised 3d models transfer to medical imaging tasks?” *ArXiv*, vol. abs/2501.11253, 2025. [Online]. Available: https://api.semanticscholar.org/CorpusID:267784726

[110] S. Azizi *et al.*, “Big self-supervised models advance medical image classification,” *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3458–3468, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:231592774